STUDY OF THE RELATIVE DOMAIN STABILITY OF A TWO-DOMAIN E.COLI MFS TRANSPORTER GLPT

Kedzierski, Mateusz Kacper

Awarding institution:
King's College London

STUDY OF THE RELATIVE DOMAIN STABILITY OF A TWO-DOMAIN *E.COLI* MFS TRANSPORTER GLPT

Mateusz K Kedzierski

A thesis submitted to King's College London for the degree of Doctor of Philosophy in the Department of Chemistry of the Faculty of Natural and Mathematical Sciences

Supervisors:

Professor Paula J. Booth

Doctor Antoni Borysik

Abstract

Members of the Major Facilitator Superfamily (MFS) make up the largest family of secondary active transporters, they mediate a diverse set of functions by controlling the movement of ions and small molecules across cell membranes. Members of the MFS share a set of common structural motifs consisting of transmembrane α-helical segments. The glycerol-3-phosphate transporter (GlpT), is an example of an MFS transporter with 12 α-helices ordered into two domains. Recent study on MFS transporter LacY, has implied an increased stability localized to the vicinity of the first helices of the protein. If this observation is found in other MFS proteins it could suggest a folding principle for other MFS transporters, whereby the first helix of the protein is acting as a stable unit that supports the process of folding. The inherent magnified stability of the helix 1 may also aid in other cellular events, where transporters or receptors are integrated into the membrane by anchoring to the membrane as well as becoming part of the unit that first penetrates the membrane leaflet. This work is focused on the stability analysis *via* alanine substitutions along the first alpha helix of the first domain of GlpT, compared to similar and corresponding mutations along the first helix of its second domain. The transporters stability is estimated by unfolding assays coupled with the decrease of secondary structure as measured by circular dichroism spectroscopy. Additional methods such as fluorescence spectroscopy, temperature denaturation and ligand binding assays have also been used in order to gain deeper understanding of the nature of the GlpT unfolding and its helical stability.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to dedicate this work to a number of people, for their support and generosity, without whom this work would not have been possible. Firstly, I would like to give special thanks to my supervisor Professor Paula Booth for giving me the opportunity to carry out my PhD research, and for all her time, guidance and feedback throughout the course of PhD study and writing of this thesis. To everyone in King's College London and University of Bristol; thank you for your friendship, and support when experiments did not work out as planned. Special thanks to Heather, Kalypso and Nicola for sharing your vast knowledge of the field and helping me to keep on top of work.

Finally, thank you to my colleagues with whom I had the pleasure of working alongside with. Thank you for you discussion and advice, this work would not have been achieved without your support and patience.

Abbreviations

Abbreviation	Meaning
ATP	Adenosine triphosphate
AU	Absorbance units
°C	Degrees Celsius
CD	Circular Dichroism
CHAPS	3-[(3-cholamidopropyl)dimethylamino]-1-propanesulfonate
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CMC	Critical Micelle Concentration
Cm	Midpoint of the denaturation curve
DDM	n-dodecyl-β-D-maltoside
DNA	Deoxyribonucleic Acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetracetic acid
g	Centrifugal Force
GuHCl	Guanidine Hydrochloride
H/D exchange	Hydrogen/Deuterium exchange
ITC	Isothermal Titration Calorimetry
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
kbp/bp	kilobase pairs or base pairs
Kd	Dissociation constant
kDa	Kilodaltons
L	Litres
LB	Luria Bertani
M	Molar
m	Gradient of the transition slope of the denaturation curve
mdeg	millidegrees
MFS	Major Facilitator Superfamily
mg	mg - milligrams
mins	minutes
ml	millilitres
MRE	Mean Residue Ellipticity
MWCO	Molecular weight cut off
OD	optical density
TB	Terrific Broth
TM	transmembrane helix
PCR	Polymerase Chain Reaction
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PS	Phosphatidylserine
rpm	Rpm - Revolutions Per Minute
RT	Room temperature
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS Polyacrylamide Gel Electrophoresis
SRCD	Synchrotron Radiation Circular Dichroism
WT	Wild Type glycerol-3-phosphate
ADP	Adenosine diphosphate
V	Volts

Chapter 1

Introduction

1.1 Introduction

Approximately 20-80 % of the membrane weight is composed out of membrane proteins (1). In total 20-30 % of all proteins in a cell reside in a membrane (2). As such, membrane proteins are responsible for the majority of plasma membrane functions. These proteins can either be directly associated with lipid tails, with the protein structure traversing the bilayer, or be anchored with the lipid heads on the surface of the membrane.

Many of the most fundamental functions required for cell function are carried out by membrane proteins. These proteins are responsible for processes such as: selective transport of molecules; resulting in a formation of a concentration gradient or electric potential; cell signalling; and receptor responding to extracellular stimuli by relaying the message into the cell. Other functions such as photosynthesis or generation of the cell's main source of energy, ATP, are accomplished by large protein complexes. These complexes are assembled from several proteins and polypeptide chains into complex molecular machinery.

Both the substrate transporters and the receptors are a major target for the pharmaceutical industry. A variety of disorders' pathologies are linked to malfunctions arising from membrane proteins. The origin of these disorders is through either the direct effect of infections or harmful mutations impacting the protein function through protein misfolding. Approximately 60 % of all drugs target membrane proteins (3), underlying the necessity of understanding the processes involving membrane proteins, including the relationship between the protein structure and function.

1.2 Protein folding studies

The process in which proteins assume their correct fold and therefore functionality is of great importance to the viability of cells. The first breakthrough in protein folding was made by Anfinsen in 1972, stating that all the information needed for correct folding of

the protein is contained within its own polypeptide sequence (4). Since then, the understanding of protein folding was expanded by multiple crystal structures and mechanisms of protein action, such as actions of chaperones and translocons in driving the protein folding *in vivo*. The technical difficulties associated with membrane protein purification and stability during experimentation have resulted in only a handful of them being studied (5). Bacteriorhodopsin is one of the most well studied membrane proteins to date. It is present in membrane patches, called the purple membrane of the *Halobacterium salinaria* organism (6). Its natural high expression and relative ease of purification, leading to high yield of pure protein, has led it to be one of the well-characterised proteins of its type. Most membrane proteins are poorly expressed and difficult to stabilise *in vitro*, resulting in complications during their study. Unstable proteins would show a tendency to protein degradation and misfolding, leading to formation of aggregates that are non-functional.

The expansion of membrane protein research would bridge the gap between the two-dimensional polypeptide and the functioning three-dimensional protein. This would result in advanced predictions of protein structure from the DNA code as well as the rational design of therapeutic compounds targeting newly discovered pathways. Moreover, the design of new enzymes and nanostructures would contribute to the expansion of the agricultural and chemical industries. Current work is focused on determining the stabilising factors of small and simple membrane proteins, so that this can later be translated to bigger and physiologically important proteins. The research in membrane protein folding is also focused on translocon-mediated folding and insertion; the role of lipids and the bilayer on the final protein structure as well as topology and protein structure prediction.

1.2.1 Structure of membrane proteins

The protein folding process is studied in a number of ways. The majority of them involve first obtaining and then analysing the folded protein structure. However, the ability to synthesize and obtain structures of membrane proteins is imperfect. In recent years, there was an increase in the number of solved structures of membrane proteins (7). Two prevailing structural motifs were identified; the α-helix and the β-barrels. The examples of two motifs can be seen in figure 1.1.

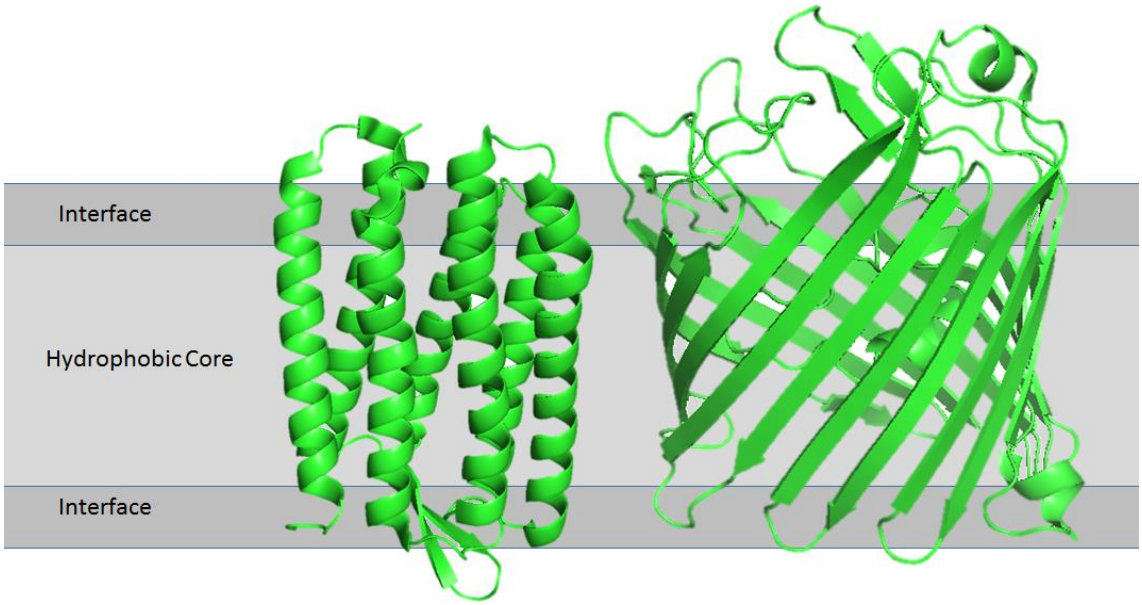


Figure 1.1 Examples of α-helical and β-barrel motifs in membrane proteins

The α-helical motif on the left is presented on bacteriorhodopsin (PDB: 4HWL, 8, Uniprot: P02945), on the right OmpF (PDB: 3HWB, 9, Uniprot: P02931) is showing a β-barrel motif. The hydrophobic core is made out of acyl chains and is approximately 30 Å in thickness. The interface, made mostly out of phospholipid head groups, is 15 Å thick. Structures were rendered in PyMOL software.

The α-helical bundles represent the dominant motif of membrane proteins, while β-barrels are mostly seen in toxins, porins and translocases. The majority of amino acids found in the membrane core are hydrophobic and are packed tightly to limit any water interactions. Salt-bridges are less common in membrane proteins than in their water-soluble counterparts. They are replaced by stabilising hydrophobic forces acting between different motifs (10). The 30 Å thickness of the hydrophobic section of

membrane dictates the average size of the secondary structure observed in the membrane. In order to accommodate the thickness of the hydrophobic proportion of the membrane the length of the α-helix or β-structure is often extended. Furthermore, α-helices can often be found to be orientated at an angle in respect to the surface plane of the membrane, with the aromatic residues such as tryptophan, located in the so-called aromatic belt, stabilising the protein through the interactions with the polar head groups (11). Other stabilising forces in the membrane proteins include: hydrogen bonds, charge-charge interactions and disulphide bonds (12). However, both the protein and the lipid membrane are constantly in motion, changing their alignment and size, resulting in a dynamic environment where the membrane protein structure constantly adapts itself (13). Through favourable interactions with lipid chains and diverse range of lipid head groups, membrane protein is able to retain both, its structure and function (14).

The protein fold and structure can also be defined by its positioning relative to the direction of the membrane. The observation that the membrane proteins tend to be found in a specific orientation to the membrane is called the positive-inside rule (15). According to the rule, certain amino acids show preference to where they are located, for example lysine and arginine tend to be found on the periplasmic side of the bacterial membranes (12, 16). It is thought that this preference helps to orientate the protein in the correct direction as well as to stabilise it during the folding process. The rule became a useful tool in predicating orientation and fold (17) of membrane proteins in the membrane.

The work described in this thesis focuses on the study of a member of the major facilitator superfamily of transporters. Thus far, all identified members of MFS were found to be composed of alpha helical bundles.

1.2.2 Alpha helical transporters

Alpha-helical bundles make up structures that facilitate multiple functions, primarily transport of substrates across the membrane. Transport is achieved by α-helices working

together in subunits to achieve substrate selectivity and *via* conformational change transit of the molecule across the membrane barrier. The substrate transport often requires a presence of the polar binding pocket, in order to induce an affinity to a substrate that is usually charged. When the transporter's channel is open, its binding pocket is exposed to an aqueous solvent. The channel structure has many requirements to satisfy in order to become functional.

The energy penalty of maintaining polar residues within the hydrophobic core of the protein needs is balanced by opposing forces in order to maintain stability and appropriate protein fold (18). At the same time the folded structure cannot be overly stable, as the transporter needs to remain flexible in order to be able to respond with the conformational change during ligand binding. The energy of transporting a molecule can be produced either from the electrochemical gradient or from hydrolysis of a molecule with high energy potential, like ATP.

Depending on the transporter, substrate transport can be facilitated in either direction. The uniporter transporters will move one substrate at a time along its concentration gradient. Symporters are able to transport different types of molecules in the same direction. An antiporter, will transport a substrate across the membrane in one direction and consequently a different molecule is ferried across the membrane in the opposite direction (19). Transporters partake in many essential physiological functions. The cell, such as neurone, will use a series of transporters with different selectivities, working in unison to provide concentration gradients of ions that then are used to drive the transport of other ions and substrates in a process referred to as action potential (20). Nearly a third of all described drug targets are thought to be transporters (21).

The structural knowledge of cell's proteome is mostly limited to soluble proteins, leaving transporters and the action of drugs acting upon them mostly unidentified. Mutations affecting the channel function or folding often lead to a disease. In the case of cystic

fibrosis, where the mutation of the cystic fibrosis gene causes misfolding of the CFTR protein, resulting in a faulty transport of chloride and in consequence thicker mucous associated with the disorder (22). The example of cystic fibrosis shows that mutations do not have to affect the binding site of the transporter in order to cause symptoms of the disease. Most often, the mutation will affect the stability of the protein through an unknown mechanism, that later would results in a disease. Understanding the processes leading to protein synthesis and stability may help to gain better understanding of possible treatment methods through identifying the relationship between the protein structure and function.

1.3 Entropy and enthalpy of folding

Multiple studies of water soluble proteins have resulted in the notion that protein folding can be described as a search of continuously more stable conformational states. As more stable contacts are favoured over non-native contacts, protein is guided towards a low energy state. The folding reaction is often depicted as a folding funnel with many local low energy states, see figure 1.2 (23).

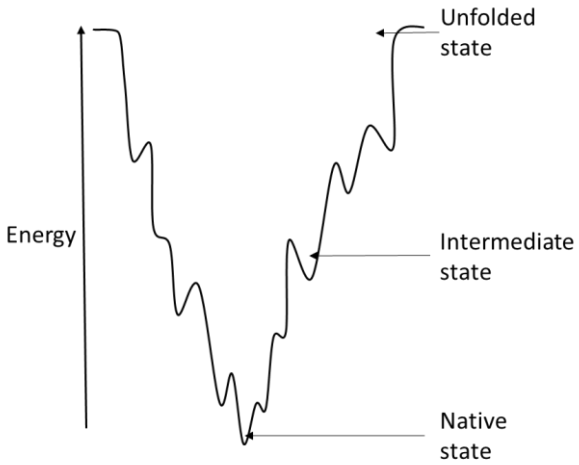


Figure 1.2 Folding energy funnel

In folding funnel hypothesis protein folding starts at the top of the funnel, with an unfolded polypeptide chain, which contains a high degree of conformational freedom and energy. The folding reaction drives polypeptide chain lower, towards protein's native state at the "bottom" of the funnel, also called the lowest energy state.

The further the folding reaction progresses towards an energy minimum the more ordered and stable conformational states are established, with the native state present at the “bottom” of the folding funnel. The shape of the folding funnel is different for all proteins. Small proteins may have a single energy barrier to pass between folded and unfolded state. More complex proteins may have multiple intermediate states, each inhabiting its own energy minima.

The hydrophobic effect is often described as the main force thought to drive protein folding. This effect is a net force, made up of multiple interactions between the polypeptide chain and surrounding environment, leading to a burial of hydrophobic side chains caused by an increase in entropy of water molecules.

For a water molecule it is more favourable to form hydrogen bonds with other molecules like itself rather than to interact with a non-polar, hydrophobic surface of a polypeptide chain. To balance for disrupting hydrogen bonds from avoiding unfavourable interactions with non-polar surfaces, water molecules form highly ordered hydration shell around hydrophobic surfaces of the protein (24). As a consequence of increasing favourable water interactions, the order of the system is increased significantly.

1.3.1 Difference between α-helical membrane protein and soluble protein folding

The folding funnel hypothesis is also used to describe membrane protein folding (7). However, unlike folding of soluble proteins, membrane protein folding is directly influenced by the properties of the lipid bilayer. During the first phases of α-helical membrane protein folding, translocon complex guides partially folded secondary structure into bilayer, where the final folding steps occur (25). Therefore, membrane protein folding starts further down the folding funnel than in soluble proteins, and is coupled with the process of membrane insertion. The final energy landscape of

membrane protein's folding funnel, and therefore the energy minima and folding rate, are dependent on interactions with the translocon and mechanical properties of the lipid bilayer.

1.3.2 The translocon mediated alpha helices insertion

Before being functional the polypeptide chain must first fold itself by forming local secondary structures and long-distance tertiary structures. The formation of structure, either in the form of α-helix or β-barrel aids in lowering of the thermodynamic cost of membrane insertion. The cost is lowered by concealing unfavourable polar interactions in the centre of the protein, thus making the protein ever more stable, by keeping these interactions away from the hydrophobic tails of the membrane's lipids (26). The initial folding steps occur within the channel of the translocon. It is thought that containing the peptide in a restricted space of the channel forces the hydrogen bonds to form interhelical connections before the newly formed helix is inserted into the membrane (27). The studies by Hessa et al. (28) with the *in vitro* expression systems provided much insight into the relationship between the character of the helix and the behaviour of the translocon. Specifically the idea of protein-lipid interactions driving the insertion of helix by the translocon. In the study, the efficiency of translocation was measured as the model helix was mutated to present a range of different amino acids. The study found that the insertion efficiency correlated with the Wimley–White whole residue hydrophobicity scale, where the hydrophobic residues were favoured by the membrane (29). Moreover, it was found that the efficiency of translocation is dependent on the position of the residue as well as its hydrophobicity. For example, the tryptophan containing helix was more likely to be inserted if the tryptophan residue was located at the end of the helix, where it would interact with the phospholipid head-groups.

The Wimley–White model describes the likelihood of observing protein residues embedded in the membrane, based on the whole residue hydrophobicity scale, which correlates well with the experimental data (30). The scale was calculated by measuring

transfer of unfolded chains from water to the POPC interface and transfer of unfolded chains into octanol. This scale can be useful in predicting the membrane spanning helix structures. However, it does not take into account other factors, such as when the helix is inserted laterally into the membrane it avoids interactions with the membrane interface (31). Moreover, different lipid compositions or stabilising effects of the tertiary structure are also not taken into the account (32).

Although, the Wimley–White model correlates well with the probability of finding residues in the membrane, protein insertion is not governed by the hydrophobicity scale alone. These scales are useful in predicting the helix structures, however they do not take into account other factors such as when the helix is inserted laterally into the membrane, where it avoids interactions with the membrane interface (33). Moreover, different lipid compositions or the stabilising effects of the tertiary structure are also not taken into the account (34).

1.3.3 Membrane protein synthesis *in vivo*

Alpha-helical membrane protein synthesis and membrane integration are thought to be primarily executed by a partnership between the translating ribosome and the translocon complex (35). These complexes are therefore linked with the *in vivo* folding steps that assist the protein to assume the correct fold and localization within the crowded and complex cell environment.

The organisation of the protein into the membrane is believed to be mediated by the degree of hydrophobicity of the peptide and the attached signal peptide (36). In eukaryotes, the process is located near the endoplasmic reticulum and executed by the heterotrimeric Sec61 complex (37). Alternatively, the bacteria possess the homologous SecYEG and accessory proteins that make up the bacterial holo-translocon supercomplex (38). Both the high resolution x-ray crystal structures and electron cryomicroscopy show much of the translocon structure at different times of the

translocation process (39, 40). The results show that to partake in the translocation process both eukaryotic and bacterial translocons undergo a tetramer formation, arranged as two dimers coupled together. However, disulphide cross-linking experiments (41) suggest that only one translocon complex facilitates the peptide translocation at one time.

During membrane protein synthesis the cell uses two pathways of the insertion by the translocon. In a co-translational pathway the ribosome with the nascent peptide chain complex displays its hydrophobic signal recognition particle (SRP) recognition sequence, spanning the length of approximately 20 amino acids at the N-terminus of the peptide (42). Once the SRP recognises the recognition sequence, the process of translation is halted until the ribosome is docked with the translocon. At the site of the translocon, the SRP receptor binds to the SRP molecule removing it from the ribosome-peptide complex and subsequently transfers the peptide into the translocon to resume peptide translation within the complex. The secretory proteins pass through the translocon to cross the membrane (43). The sequence that identifies the proteins as membrane bound is unknown, but it is thought to be associated with the degree of hydrophobicity.

An alternative post-translational process starts with already translated protein. In bacteria, newly translated proteins are coupled to SecB and delivered to SecA chaperone. Using ATP hydrolysis, SecA transfers the protein through the translocon using ATP-mediated conformational changes (44). The majority of secreted proteins are directed through the post-translational pathway. In eukaryotes the Sec62/63 complex is used to couple the protein with the translocon. Before being functional the polypeptide chain must first become folded by forming local secondary structures and long-distance tertiary structures that are arranged in the membrane and are described as the protein's fold or topology.

1.3.4 Membrane protein topology

Before the protein assumes its final structure the helices need to arrange themselves in the correct orientation in relation to the membrane. The translocon is thought to take part in determining protein topology, when the peptide interacts with the lateral gate of the translocon complex. Subsequently, when the membrane helices are inserted into the membrane the final folding step occurs, and the protein assumes its native tertiary contacts.

The information that decides the membrane protein topology is coded in the amino acid sequences and depends on a number of factors (45, 46). Amino acid residues bearing a positive charge, according to observations of the positive-inside rule (15), are more frequently found on the cytoplasmic side of the membrane. It is possible that these residues are not only guided by the proton gradient across the membrane, but also interact with the translocon itself (47). The scale of residue hydrophobicity is also a key factor in determining the topology. The hydrophobic residues will find their suitable positions within the segments of the membrane so that the net thermodynamic cost of inserting helices is satisfied. Moreover, the lipid structure and membrane composition will act as key determinants, which fine tune the protein structure during and after the insertion (48, 49). Post or co-translational modifications, such as glycosylation, may also influence protein topology (50).

In the end, the structure of the helices may depend on interactions with other neighbouring helices (51). The general belief that transmembrane helices are inserted into the membrane in the same sequential order as they are translated by the ribosome could only be applicable to few proteins. The cross-linking studies of aquaporin 4 insertion (52) have shown that some helices would be first integrated into the membrane only to return to the translocon to interact with the downstream helix that just entered the complex.

1.3.5 Models of protein folding

The problem of biological macromolecule folding is thought to be one of the most difficult challenges of modern biophysical research. Among thousands of possible conformations proteins are able to assume a unique stable structure in a shorter time than that which is required to test all possible orientations by trial and error (53). Therefore, it is likely that protein folding is governed by a set of rules and restrictions that guide the entire process. Many models were proposed to describe the process of folding α-helices within the membrane.

The two step model, developed by Popot and Engleman in 1990 (54), states that first, helices are formed as independent, self-assembling units across the lipid membrane, followed by inter-helical interactions that form the tertiary structure. The observations of bacteriorhodopsin being able to refold from the independently stable units seem to confirm the properties of a single helix acting as an independent unit (55, 56). Although the model does not include additional phenomena such as cooperative folding in large multi-domain proteins, it does provide a fundamental principle that can be used to develop other more complex models, which take additional phenomena into account.

The third stage to a two-step model was introduced by Jacobs and White in 2003, to incorporate additional events occurring after the helices self-associate. The folding of certain proteins may involve extra steps such as: binding of cofactors, folding of loops and insertion of extra polypeptides into the membrane (57). The extra events add complexity to the three stage model, however structural studies have observed the presence of many folds that cannot be explained by a two or three stage model. In order to address these observations, the four stage thermodynamic model (26) was introduced. It represents the combined arguments of Popot-Engelman's two stage and Jacobs-White three step models to describe protein folding in terms of:

-Insertion of the unfolded peptide into the membrane interface;

-Folding of the helix at the interface;

-Insertion of the pre-folded helix;

-The association of multiple helices into tertiary structure.

Like other models, the four step model does not account for cascades of smaller steps that make up the folding process nor the final collapse, which finalises the folding of the secondary and tertiary structures (58, 59).

1.3.6 Common motifs in membrane protein structure

Common motifs and peptide sequences have been identified in order to compose a set of features that aid in maintaining protein stability. Many of these motifs are only applicable to few proteins, while others are more wide spread, such as the stretches of subsequent hydrophobic residues that were proven to be good indicators of transmembrane domains. Other motifs like, the GxxxG motif, were identified by Russ and Engelman (60) as the most common arrangement of residues in transmembrane helices. The motif describes spacings of glycine (GxxxG) and alanine (the AxxxA motif), resulting in an arrangement of small residues on a single side of the helix. Over 50% of α-helices displaying the GxxxG motif participate in helix-helix interactions (61). The small residues create cavities into which the larger residues can fit, increasing the compactness of the protein as well as increasing protein stability through the exclusion of water and promoting the Van der Waal interactions. This ‘’knobs-into-holes’’ concept encourages strong helix-helix interactions, but has also been implicated in the function and targeting of transmembrane proteins (62, 63).

The membrane spanning sequences are dominantly hydrophobic. However, the occurrence of polar resides, such as asparagine, through which strong intermolecular bonding can be used to drive inter-helical formations (64), as well as to participate in water interactions. Aromatic compounds, such as tryptophan and tyrosine, can also be used to a similar extent to elicit helix oligomerisation, as well as promote interactions with the lipid head groups found in the bilayer (65, 66).

Although the recognition of common motifs, such as amino acid sequence, may be used in prediction of the secondary and tertiary structures, the lipid membrane also needs to be taken into account as a major factor which influences protein structures. The property of the lipids within the membrane could be used to locally favour one conformation over another. For example, the short chain lipids may force the helix to tilt in order to bury its hydrophobic residues in the centre of membrane.

1.4 Membrane composition

The two basic components of biological cell membranes are lipids and proteins. The lipids maintain one of the most basic functions, which is to act as a barrier between the environment surrounding the cell and the cell's organelles. The difference between the two environments allows for the formation of electric and chemical concentration gradients that drive many of the functions of the cell. Furthermore, membranes are involved in cell communication. They facilitate cell motility as well as provide surface for adhesion of other protein structural components, like cytoskeleton and extracellular matrix (67). A diagram of the biological cell membrane can be seen in figure 1.3.

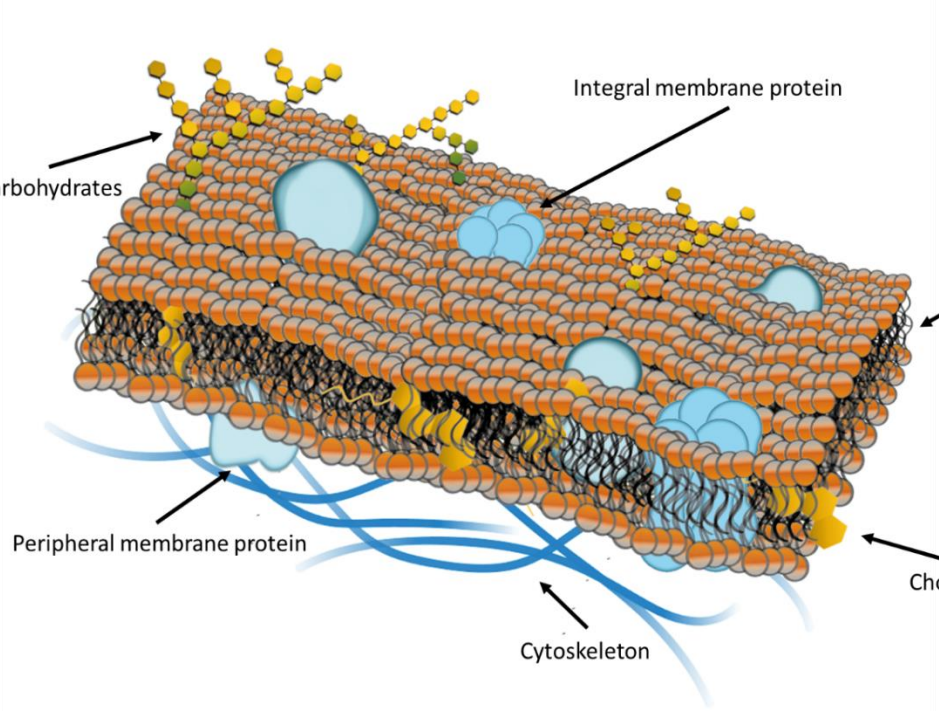


Figure 1.3 Schematic of biological membrane

The fluid mosaic model of a cell membrane presents lipids self-assembling into two leaflets with hydrophobic lipid tails facing inwards. Proteins can be found either within the membrane (integral proteins) or associated with the membrane surface (peripheral proteins). The cytoskeleton links the membrane to other cell organelles. Cholesterol is found in the centre of the membrane, contributing to the membranes physical properties. Carbohydrates are found bound to lipids (glycolipids) or proteins (glycoproteins) participating in cell adhesion and tagging. The figure is adapted from (68).

1.4.1 Membrane lipids

The diversity of membrane function is mirrored by the diversity of its constituents. A number of lipids make up the bilayer. The three major types of lipids are phospholipids, glycolipids and sterols (69). Phospholipids form the majority of all cell membranes. Their structure varies by their head groups as well as length and saturation of their acyl chains. The composition and the structural variety of lipids confer the membrane properties such as the degree of elasticity, fluidity and lateral pressure (70). The phospholipids can be found with one or more carbohydrate groups attached to the head group in the outer leaflet of the bilayer. Cholesterol, present in only eukaryotic cell membranes, maintains

the structure and fluidity of the membrane and is present to a varying degree in tissues of all animal cells (71). The function of lipids is not limited to providing structure to the cell membrane, but they also may act as an energy storage and act in cell signalling (72).

The lipid bilayer is a very dynamic part of the cell that constantly adapts itself in response to the outside environment and the presence of membrane proteins (73). The most common types of lipids found in eukaryotic cell membranes are phosphatidylserine (PS), phosphatidylcholine (PC), phosphatidylethanolamines (PE) and sphingomyelin (SM), as well cholesterol. The chemical structures of various membrane constituents can be seen in figure 1.4. In comparison, the cholesterol is present in only small amounts in the membranes of mitochondria and endoplasmic reticulum organelles (71). More primordial cell membranes, such as in gram negative bacteria *E.coli*, are rich with PE, but lack cholesterol. Furthermore, the inner membrane of gram negative bacteria is mostly made out of PE with the addition of phosphatidylglycerol (PG) and cardiolipin, whereas the outer membrane is composed mainly of lipopolysaccharides.

Lipid self-assembly in an aqueous environment is driven by the amphipathic character of the lipid molecules. The hydrophobic effect compels the acyl chains of lipids to come together, while exposing the hydrophilic head groups. The structures that satisfy the hydrophobic effect reduce their unfavourable contacts between the acyl chains and water by assuming one of the different mesophases, such as: lamellar, cubic, hexagonal, and micellar. Both the environment and the chemical structure of the lipids dictate into what phase they will self-assemble. The state of the lamellar phase or its fluidity depends on the temperature and is influenced by the acyl chain saturation and length as well as the type of the polar head group. As the temperature decreases the fluid phase changes first to a gel phase and then to a liquid crystalline phase (71). As the hydrophobic forces keep both layers flat in the bilayer, the curvature-increasing lipids tend to increase the lateral pressure within the membrane (74).

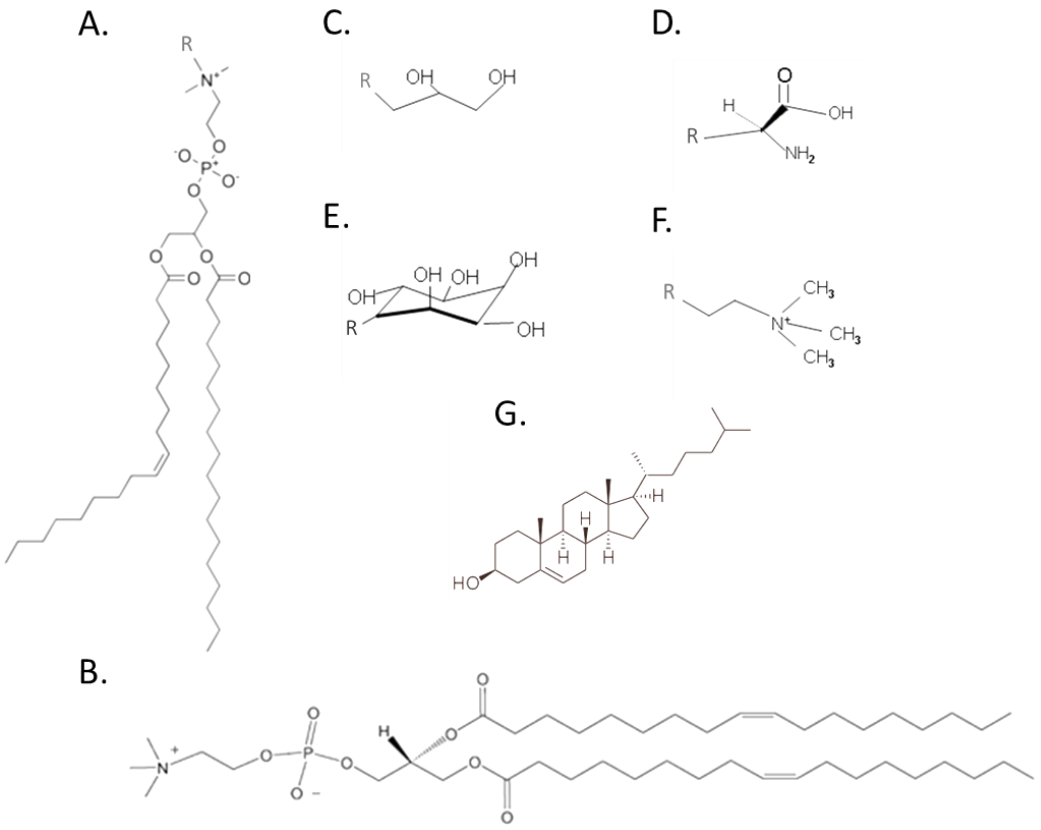


Figure 1.4 Structures of phospholipids and cholesterol

The structure of an un-saturated phospholipid, PE, is seen in panel A. The alternative lipid without a kink with saturated acyl chain is shown in B. The four common phospholipid headgroups are listed as C – phosphatidylglycerol (PG), D – phosphatidylserine (PS) E – phosphatidylinositol (PI) and F – phosphatidylcholine (PC). The component of the lipid bilayer, cholesterol can be seen in panel G. Figures were created in Chemdraw software PerkinElmer Informatics (http://www.cambridesoft.com).

1.4.2 Studying membrane proteins *in vitro*

A growing number of *in vitro* studies concerned with membrane proteins seek to mimic the *in vivo* processes and report the key influence of the lipid composition on the protein behaviour and structure (75 - 77). For example, examination of PE-deficient *E.coli* strains have shown that the orientation of lactose permease transporter (LacY) is PE dependent. When PE is absent, the N-terminal part of the protein was found inverted. Re-introduction of PE into the membrane restored the correct orientation of LacY (78). Consequently, when describing the properties of membrane proteins, it is beneficial to include the

effects of the membrane on the studied protein, since both the structure and behaviour of the protein can be heavily dependent on the composition of the membrane.

The complexity of the plasma membrane environment is difficult to reproduce and makes the biophysical measurements very difficult. Several biomimetic systems exist that allow studying of proteins in *in vitro* systems; these include lipid monolayers, bilayers and lipid vesicles (79). The fundamental property of these systems is the solubilising unit that enables the protein to assume its native structure. Later, the composition of the system is modified by addition of more solubilising components or other folding co-factors. Once the membrane protein is isolated and solubilised it often becomes unstable and tends to aggregate in the solution. A variety of solubilisation methods and chemicals were developed to address the issues of membrane protein solubilisation and the ability of the experimental *in vitro* system to mimic the environment of the cell.

1.4.3 Detergents and lipid systems

Detergents are amphiphilic chemicals that are commonly used in membrane protein purification. They are composed of both hydrophilic and hydrophobic groups that together close the gap between the two different environments. The polar head group interacts with the aqueous environment, while the hydrophobic tails orient themselves towards apolar conditions. In an aqueous environment the detergent molecules bury their hydrophobic tails by assembling into micelles. Micelle formation is concentration dependent and to successfully observe micelles the concentration of detergent needs to be kept above the critical micelle concentration (CMC), which is specific for each different detergent. Below the CMC, the detergent molecules are still soluble, but persist as monomers. To solubilise the membrane protein a concentration greater than the detergent's CMC is used (80). The micelle then forms around the protein mediating the interactions between the protein's hydrophobic regions and the polar aqueous environment.

The choice of the right detergent can be important, as the conditions of the purification and the experiment may differ. The right balance needs to be found to choose a detergent in which the protein is stable, while at the same time remains active (81). Certain detergents force the protein to adopt a new conformation that is similar to the native fold, but through either formation of new contacts or occlusion of the active site, renders it unable to retain functionality (82). Both size and shape of the micelle depends on the detergent forming it. In the detergent-protein complex, the micelle will change its shape and size in order to protect hydrophobic stretches of the protein from aqueous solution.

Detergents can be classified into four groups based on their structure. The ionic detergents, such as sodium dodecyl sulphate (SDS), are excellent at solubilising the protein, but at the same time often result in denaturing the protein to some extent. The SDS-denatured protein can sometimes be refolded by a transfer to a more favourable detergent system. Bile acid salts, such as sodium deoxycholate, are structurally related to cholesterol and are considered as mild detergents, seldom deactivating the protein (80). Nonionic detergents, DDM, are mild and non-denaturing, however short chain hydrophobic tails tend to result in a loss of protein functionality. Zwitterionic detergents, such as CHAPS, combine both positive and negative charges. They are favoured for structural studies, as they are less deactivating than nonionic detergents (80). Examples of chemical structures of various detergents can be seen in figure 1.5.

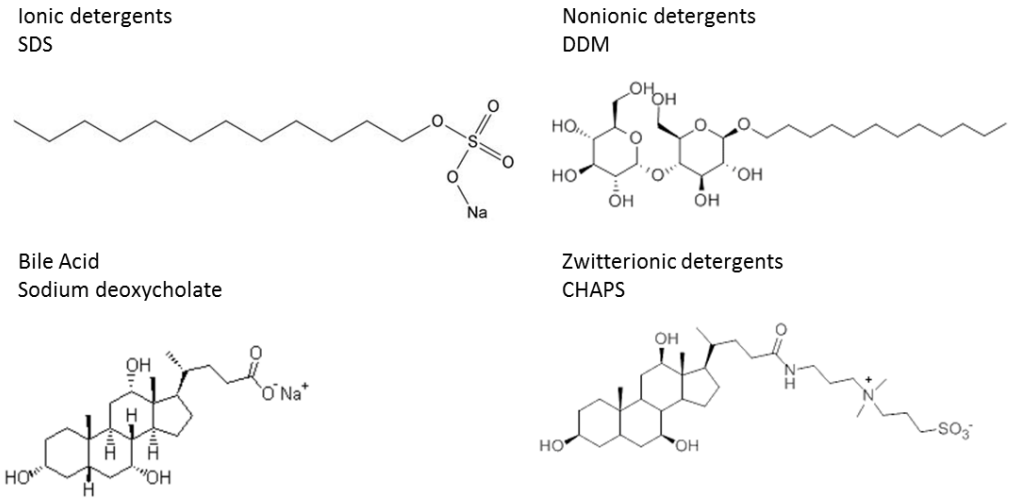


Figure 1.5 Comparison of detergent structures

The structures of four different types of detergents: ionic, nonionic, zwitterionic and bile acids are shown above. Figures created in Chemdraw software PerkinElmer Informatics (http://www.cambridesoft.com).

Although detergents are successful at solubilising membrane proteins and aiding in their purification they do not mimic the complexity of the plasma membrane. However, certain structural studies require the system to remain simple.

The importance of studying membrane proteins is reflected by the focus on developing solubilising systems that mimic some of the membranes properties. One such type of systems are nanodisc structures, which utilise a small lipid bilayer, surrounded by the scaffolding protein ‘’belt’’ that protects the lipid chains (83). A new system employing styrene–maleic acid copolymers (SMAs) instead of scaffolding proteins was used to purify membrane proteins directly from their lipid environment in soluble nanodiscs (84). Alternatively, the proteins can be solubilised using amphipathic polymers called amphipols (85, 86).

The ever growing diversity and size of the artificial membrane protein solubilisation systems shows the intrinsic tolerance of membrane proteins to the chemical character of solubilising agents. Whether it is *via* detergent, peptide or native lipid, the membrane protein is able to maintain the majority of its secondary contacts, as the results from different model systems can be similar to each other (87, 88). Moreover, this tolerance is preserved across different domains of life as an evolutionary advantage that allows organisms to alter their lipid composition of membranes according to the environment they are challenged with (6). However, the importance of lipid-protein interactions cannot be ignored, as the native membrane environment is used to modulate protein functionality and location through dynamic cell-specific variations.

1.5 Major facilitator superfamily

Most studies are focused on small, single domain proteins that are relatively easy to purify and study (79). Focusing the studies on a single family gives the benefit of uncovering some shared principles behind membrane protein folding. Moreover, studying the large multi-domain proteins gives the opportunity of investigating interactions between different domains and their shared contribution to the stability of the protein. In the case of the family of transporters, the relationship between these domains also drives the transportation function of these proteins, giving an insight into the function-fold relationship.

The major facilitator superfamily (MFS) is one of the largest known families of structurally related secondary transporters observed both in eukaryotes and prokaryotes. A quarter of all transporters within the cell are in the major facilitator superfamily (89). They are known to facilitate the transport of a wide range of substrates, including: ions, drugs, sugar phosphates, amino acids, peptides and nucleosides (90). In this thesis a member of the MFS family, GlpT, is investigated. Its stability and functionality, as well as the influence of the two domains on the protein structure are discussed and compared to the

other known and studied members of the same family. The two examples of the *E.coli* MFS transporters, whose stability was previously investigated, are the lactose permease (LacY, 91) and galactose transporter (GalP, 92). The answer as to whether or not these domains exist and fold independently of each other or rather cooperatively fold and assemble *in vivo* will also help in shaping the understanding of their mechanism of transportation. Most of the previous research into MFS transporters was related to their substrate selectivity and mode of transport. However, little is known about the MFS conformational stability *in vitro*.

1.5.1 MFS structure

Regardless of low sequence similarity, structures of all resolved MFS transporters share a similar structure of 12 transmembrane α-helices, which are split into 2 domains, with an aqueous cavity located between the two domains (93). It is also believed that most MFS transporters share the same rocker-type switch mechanism of transport (94). Some transporters from the MFS family have an additional 2 transmembrane helices, for example transporter PepT. The function of the extra helices is not known, but is thought to serve functions other than transport (95).

The first structure of an MFS transporter to be resolved was lactose permease (LacY) (96) in 2003. This was soon followed by the structure of the glycerol-3-phosphate transporter GlpT (97), and by the multidrug transporter EmrD (98), the fructose transporter FucP (99) and the xylose transporter XylE (100). Since 2011, 15 more structures of MFS transporters were solved, those include human glucose transporter GLUT1 (101, 102). All of the resolved structures show 12 transmembrane α-helices arranged in characteristic two domains, enveloping the central cavity responsible for transport. Examples of MFS transporters structures can be seen in the figure 1.6.

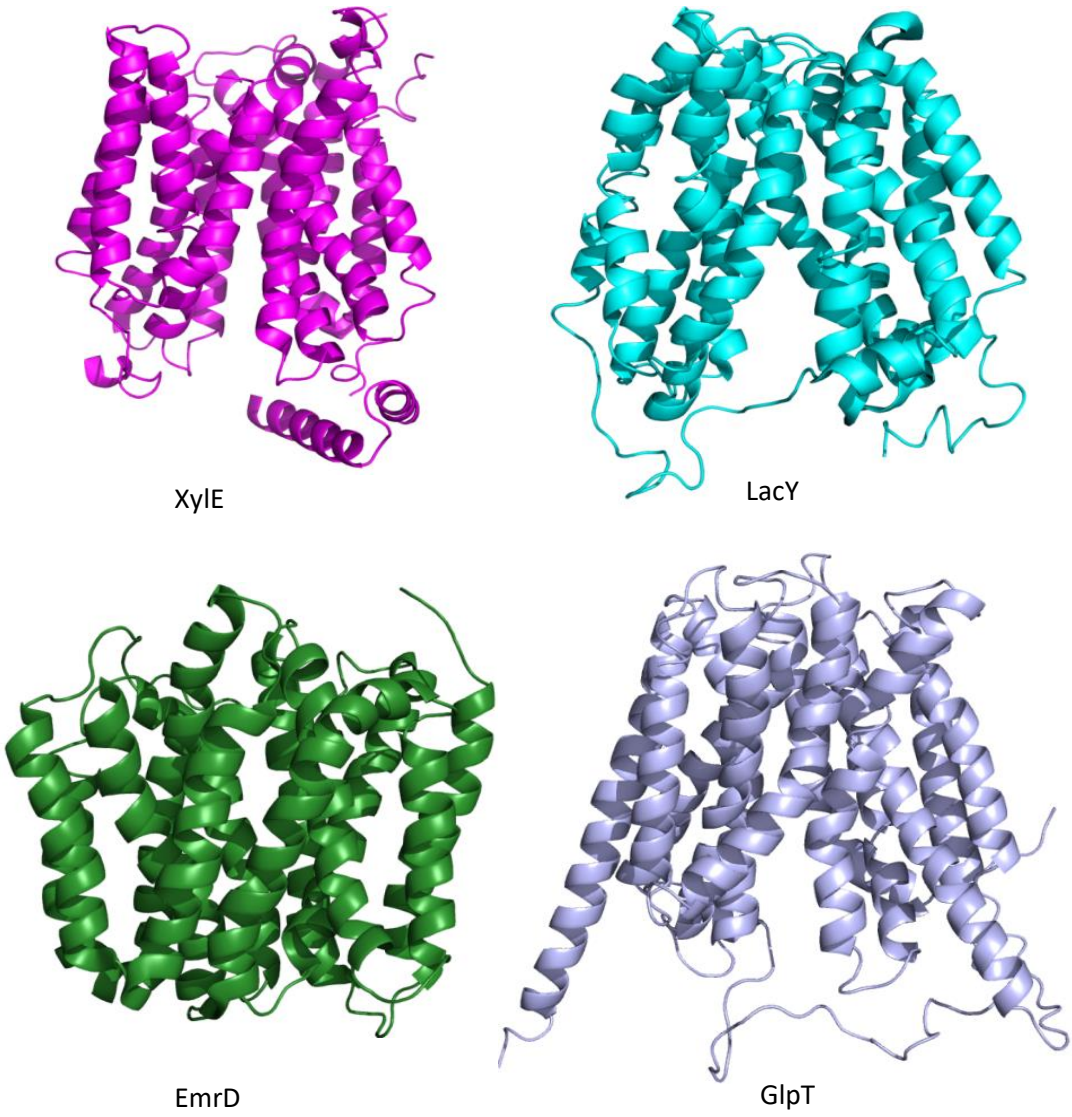


Figure 1.6 Examples of researched members of the MFS transporters

XylE, above, is in an outward-facing and partially occluded conformation (4GBZ, 100, Uniprot: P0AGF4, 23% identity and 39% similarity, 4.11 RMSD, 1110.14 Z-score with GlpT), LacY is in an inward facing conformation (2CFQ, 104, Uniprot: P02920, 20% identity and 38% similarity, 3.80 RMSD, 1300.66 Z-score with GlpT), EmrD is in an intermediate state (2GFP, 98, Uniprot: P31442, 23% identity and 37% similarity, 3.57 RMSD, 335.42 Z-score with GlpT) and GlpT is in an inward facing conformation (1PW4, 97, Uniprot: P08194). Structures were rendered in PyMOL software. Analysis of sequence alignment (similarity and identity) was done in RCSB PDB Protein Comparison Tool (105) with Needleman-Wunsch sequence alignment method (106). Analysis of structure alignment (RMSD and Z-score was also performed in RCSB PDB Protein Comparison Tool using jCE algorithm (107).

The genetic sequence research (95) provided evidence of the MFS structure origin. It is thought that the MFS structure similarity did not arise from an early triplication event of two transmembrane helices, but rather the duplication of three helices, which gave rise to six helical bundles. Through an additional gene duplication event, two six helical bundles gave the MFS characteristic structure of six helices arranged in two structurally similar symmetrical domains. According to the IUBMB-approved Transporter Classification Database (103), there are 74 distinct major facilitator transporter families within the superfamily. In addition, each transporter has its own set of substrates. Out of the 74 families, 17 are not functionally characterized yet. The diversity of substrates transported by MFS is thought to have arisen from a common precursor from which additional substrate selectiveness was incorporated through mutations. Subsequent genetic events, resulting in frequent domain swapping between different members of the MFS families, have led to the substrate diversity currently observed within these transporters.

The diversity of a family as big as MFS relates to their importance in viability of the cell. It is therefore not surprising that the disorders targeting MFS transporters were found to contribute to several diseases. The examples include an organic anion transporter causing renal failure through harmful accumulation of indoxyl sulphate (108), mutations in peptide transporter (PepT) leading to inflammatory bowel disease (109) and glucose transporters (GLUT family) associated with type 2 diabetes and cancer (110). Unfortunately, the technical issues associated with studying membrane proteins have made characterisation of many disease-implicated transporters unattainable. Furthermore, the ubiquitous expression of mammalian MFS proteins is further complicated by an inherent instability of these proteins in *in vitro* conditions, as well as the need for post-translational modifications required for their function. The already mentioned conservation of structure seen in MFS transporters allows for predictions of the unresolved mammalian structures by choosing a close bacterial homologue template of known structure in order to identify possible protein folds through homology modelling.

Both LacY and GlpT were used many times as model transporters to predict the structure of physiologically important mammalian proteins, such as the glucose 6-phosphate transporter, G6PT (111), or GLUT1 (98, 112)

The function, variety and size of MFS transporters highlights the importance of their contribution to mammalian physiology. Studies that improve the understanding of the MFS proteins help to uncover the identity of unclassified transporters, as well as their input into debilitating conditions. Moreover, expanding the folding and stability knowledge of similar proteins within the same family has the benefit of uncovering shared behaviour and trends that contribute to their stability and folding pathway (113).

1.5.2 MFS mechanism of transport

All members of the MFS family are thought to share a common mechanism of transport based on the high degree of shared structural homology. The most supported model for an MFS transport is through the mechanism of alternating access, also called the rocker-switch (114). This mechanism is based on structural conformations induced by ligand binding. Diagram of the mechanism can be seen in figure 1.7.

First, the aqueous binding pocket is exposed to one side of the membrane, allowing the substrate to interact with the residues located in the binding pocket. Ligand binding induces the shift of the two domains, leading to two events. The previously open side of the pore closes with hydrophobic residues blocking the substrate withdrawal. At the same time, the second side of the membrane opens, exposing the binding pocket to the aqueous media (115). The inversion of the protein structure during the rocker-switch cycle involves breaking and formation of several salt bridges as shown in LacY (116) and GlpT (117). However, as the crystal structures show, the arrangement of helices remains mainly undisturbed (118). The ligand is released, the transporter returns to its original confirmation and the cycle is repeated. The main support of the alternating access model comes from crystal structures of the transporters themselves. Further

evidence comes from the deactivation (117, 119, and 120) and cross-linking mutation studies (121), molecular dynamics simulations (94, 117, and 122) and residue labelling experiments (123).

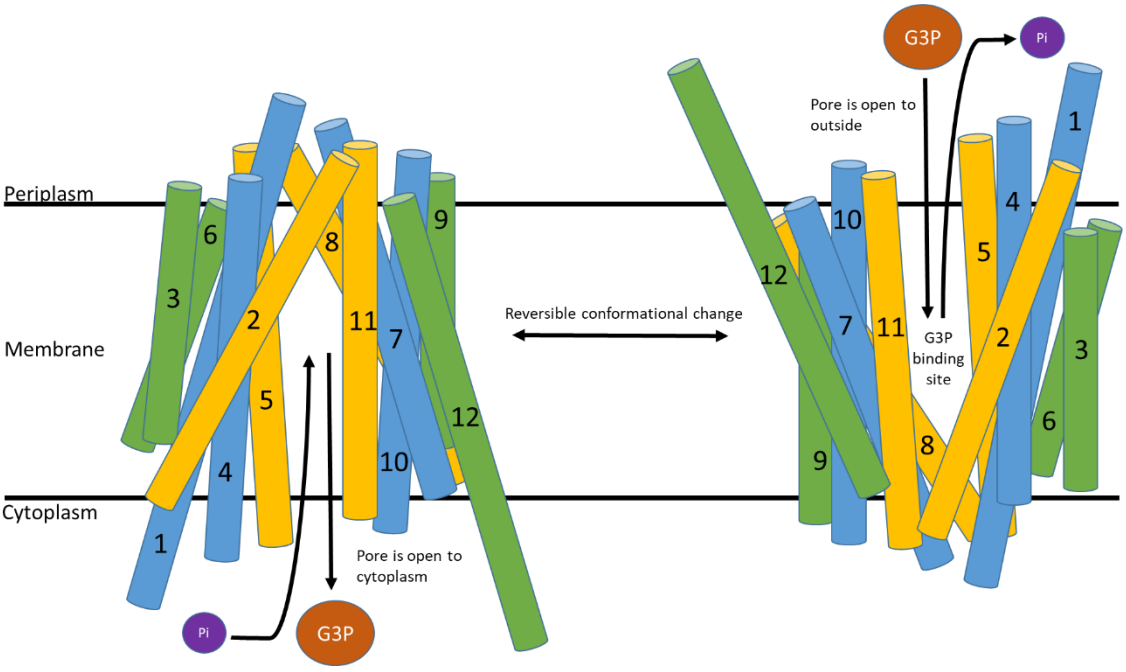


Figure 1.7 Schematic representation of GlpT rocker-switch mechanism of transport.

The transport cycle shows inward-facing and outward-facing conformations of GlpT as it transports a G3P molecule from periplasmic environment to cytoplasm in exchange for a phosphate molecule. The mechanism involves a reversible conformational switch occurring upon G3P binding leading to inversion of the protein structure. The helices are numbered sequentially from the N-terminal end.

1.5.3 GalP

GalP is a proton-galactose transporter in the MFS and its expression is readily induced in *E.coli*. Both equilibrium unfolding and refolding curves were established in the DDM-solubilised protein with the free energy of the folding reaction calculated (124). The protein refolding was assayed by both return of the secondary structure and the return of effective substrate binding. Approximately one third of the protein structure was lost when the protein was denatured with 8 M urea. The kinetic measurements of GalP unfolding have shown a single exponential trend of unfolding, supporting the two state

unfolding mechanism. Moreover, the GalP stability was investigated under varying lipid compositions, which showed that refolding was more efficient when reconstituted directly into lipid vesicles, rather than in DDM micelles. However, the return of the helicity was not related to the return of the original transport activity, compared with the folded reconstituted protein. This suggests the helical packing not returning to the native state is subsequently affecting the transport efficiency.

1.5.4 LacY

LacY is one of the most studied members of the MFS. It facilitates the transport of galactopyranosides by using a negative proton gradient. The structure of LacY was successfully unfolded and refolded in the DDM micelles, with binding of the NPG (p-nitrophenyl α-d-galactopyranoside) ligand returning upon refolding, measured *via* ITC and the return of secondary structure, measured by far-UV CD (125). The structure of LacY was shown to be very dynamic with a fast rate of hydrogen\deuterium exchange achieving 90% of labelling in less than 5 minutes (96). The high rate and extent of deuterium labelling could also be an indication of unfolding process in the sample. LacY has a high tolerance to mutations with only 6 residues being irreplaceable for the successful transport function. An experiment in which, a series of single-Trp mutants of LacY were analysed in a urea denaturing assay, has found an increased stability of the first helix of the LacY's first domain, compared to the rest of the protein. It was therefore proposed that a stable helix 1 could aid in the process of membrane insertion (126).

1.5.5 GlpT

The glycerol-3-phosphate transporter is a novel protein to include in the studies of the MFS transporters. Compared to the LacY and GalP, GlpT is a much larger transporter that facilitates transport *via* an exchange of two substrates. The structural similarities between these transporters help to draw attention to a common folding pathway and sources of potential stability. Glycerol-3-phosphate transporter (GlpT) shares 90% of

sequence similarity with LacY. Stability and composition of the GlpT and its helices has already been studied *via* molecular dynamics simulations and mechanical denaturation (127, 128) giving context to the chemical stability measurements discussed in this thesis.

Homologue transporters related to GlpT are widely spread and are found in bacteria, eukaryotes, plants and humans (129). Several studies on the function of MFS transporters have highlighted GlpT as a structural and functional model of the MFS family, along with LacY (59, 94, 118 - 122). The common topology of MFS transporters is also found in GlpT. The two domains are only limited in their relative movement to a few hydrogen bonds that exist in the interface between the domains made by helices 1 and 7 (97). The overall native structure of GlpT is therefore easily adaptable and flexible enough to facilitate transport *via* alternating access.

1.5.6 GlpT binding and function

GlpT is located in the inner membrane of *E.coli,* where it facilitates uptake of glycerol-3-phosphate (G3P) in exchange for inorganic phosphate (Pi). G3P acts as a substrate for glycerophospholipids, the main component of the membrane. (119). In the absence of other substrates, G3P may serve as a source of carbon and energy for *E.coli* (130). It is thought that Pi binding on the cytoplasmic side induces the conformational change that opens the transporter to the periplasmic side. Replacement of Pi with G3P, on the periplasmic side, causes the transporter to revert to the original cytoplasm open state. In the absence of G3P, GlpT is still able to mediate exchange of Pi across the membrane (117). The two arginine residues, R45 and R269, which line the central pore, display a strong positive charge, which mediates binding to the GlpT substrates. Additional residues K80 and H165 have also been found to have a role in stabilising the phosphate moiety of G3P or Pi (119, 131). Residues Y38, Y42, Y76, Y266 and Y393 are also involved in transport, but do not interact with ligands directly. Instead, they are important in closing the binding pocket during the alternating access transport.

1.5.7 GlpT in disease

The G3P molecule was found to take part in metabolic control of low density lipoprotein. It is also an important precursor to lipid synthesis in eukaryotes (130). In the glycerol phosphate shuttle system, G3P is used to protonate NAD+ at the inner membrane of mitochondria (132). The SLC37A1 protein is believed to be a human equivalent of glycerol-3-phosphate transporter (133). The SLC37A1 gene is widely expressed in all tissues, with a higher rate of expression in kidney, bone marrow and liver. It has also been found to be upregulated in breast cancer and endometrial cancer cells. An increased phospholipid biosynthesis due to SLC37A1 gene upregulation in cancer was proposed to drive tumour proliferation (134). Furthermore, GlpT, along with the hexose phosphate transporter (UhpT), was found to facilitate transport of antibiotic phosphomycin (135). The inactivation of the GlpT gene results in a significant increase in phosphomycin resistance in *Pseudomonas aeruginosa,* which is responsible for nosocomial outbreaks in cystic fibrosis (22).

1.6 Differences in helix stability

Transporters within the MFS family share topology of 12 transmembrane helices found between two domains. Studies performed on LacY (125) suggest that, certain protein regions show increased stability and may also possess a functional role in the process of folding of the MFS proteins. It is believed that the structure of the first helix of the first domain shows increased stability when compared to the remainder of the protein (136, 137). The process that leads unfolded protein towards the folded and functional structure is driven by forces that influence the interactions between amino acids towards the lower energy state. In order to probe the individual contributions of amino acids in the folding process, an alanine scan methodology is implemented and discussed in this work (138).

The measured change in the free energy upon the substitution mutation is composed of several components and can be interpreted as missing covalent interactions,

rearrangement of hydrogen bonds, changes in solvation energy, non-covalent interactions and changes in any additional energy terms caused by rearrangement of protein structure (139). The experiments described in chapter 4 are designed around the alanine scan methodology and are based on the similarity of environments between two chosen mutation sites on two similar helices, in order to cancel any extra energetic terms and make the analysis of energy differences simpler. Each mutation on helix 1 will have its corresponding second mutation located on helix 7, the first helix of the second domain. Like in all MFS proteins, the two domains structurally resemble each other. This second mutation will be selected based on a similar amino acid substitution change as well as a similar electrostatic environment to its respective pair. As a result, both helices will be subjected to a similar series of mutations, where possible, based on the mutation position and immediate environment of mutated residue.

1.6.1 Alanine substitutions

Systematic substitutions of amino acids to alanine residues are used to estimate the individual contributions of the targeted residues to the protein stability and functionality (140). Although alanine scanning is employed widely in the analysis of soluble proteins, membrane proteins have lagged behind due to their low expression and complications arising from methods of solubilisation. Moreover, the final membrane protein structure and function are affected by the micelle environment, as well as alanine substitutions. Therefore, the use of alanine substitution method requires understanding how a substitution mutation has affected the structure of protein and protein-micelle complex. However, this method can be used to quickly generate a series of mutations, targeting protein regions of interest. Unless, combined with another method, the results are missing any structural information and only relate to general change of protein function or stability. The method is most useful when applied in a high-throughput screen. In this approach, through targeting and elimination of molecular contacts, functional binding sites for molecular interactions can be found, as well as contributions of amino acids to protein stability.

Frequently, studies employing alanine mutations have focused on site-specific, non-conservative substitutions of charged amino acids in order to assay the role of these residues in the protein function. An example of alanine scanning methodology can be seen in a systematic substitution of charged residues on the surface of the epithelial sodium channel (142). The mutations identified sites that are responsible for modulating channel conductance and are also essential for the transport of proteins to the cell surface. Other examples involve the use of alanine mutations to determine specificity of α3β4 nicotinic acetylcholine receptor (143), determination of functional epitope of phosphatase type 1 in *Saccharomyces cerevisiae* (144), contributions of the salt bridges to activity of *Pseudomonas aeruginosa* elastase (145), as well as in protein folding (126, 137). The final analysis step is time consuming, as each alanine mutation needs to be assayed for potential change in the protein fold due to the mutation. Furthermore the folding pathway needs to be investigated by the unfolding and refolding assays.

1.7 Aims of the project

A recent study on the MFS transporter LacY (125) has implied an increased stability localized to the vicinity of the first helix of the protein. If this trend is extended to other membrane proteins it could suggest a folding characteristic of the integral membrane proteins, where the first helix may aid in cellular events, such as integration of the protein into the membrane.

The aim of this study is to examine the difference in stability between different helices of GlpT. This will be done by identifying whether or not the first helix shares the same stability as the first helix of the second domain, helix 7. The folding of GlpT is investigated by the combined measurements of fluorescence and circular dichroism spectroscopies during urea, GuHCl and thermal denaturation. The available crystal structure of GlpT (1PW4, 97, Uniprot: P08194) is used to discuss and put the results into context.

The urea-induced unfolding was found to be reversible in both GalP and LacY. Chapter 3 will assess the extent of GlpT unfolding and its ability to refold, as well as the development of the substrate binding assay that will later determine the return of transporter functionality upon refolding of the chemically denatured state. Chapter 3 also focuses on the single-Trp variants of GlpT, as well as the design and function of the double rhodamine labelled GlpT and whether they can be used as molecular reporters of unfolding.

As the suitability of GlpT to the unfolding study is established, the main hypothesis of the GlpT's helix 1 stability is approached in chapter 4. The main narrative of chapter 4 is focused around the consecutive alanine substitutions of amino acids located on helix 1 and 7. The effect of each mutation on stability is then compared against WT and similar mutations on helix 7. As a result, both helices are subjected to a similar series of mutations, based on the mutation position and immediate environment of mutated residue. Any differences in stability between related helices would highlight complex environment in this membrane protein, as well as show likeness to another protein, LacY in the same MFS superfamily.

This thesis will show that through multi-approach techniques of probing the protein behavior it is possible to gain further insight into the stability of a multi-domain membrane proteins. Use of single alanine substitution analysis in measuring the differences in unfolding of different parts of GlpT will be discussed. The importance of binding pocket and domain interface in contributing to protein stability will also be revealed.

Chapter 2

Materials and methods

2.1 Reagents

The majority of the reagents used were supplied by either Sigma-Aldrich or Fisher Scientific UK Ltd, with the exception of the following:

Detergent DDM (*n*-Dodecyl-β-D-maltopyranoside) obtained from Anatrace, Amicon spin concentrators obtained from Millipore, mGelRed, obtained from Biotium and The Quick Coomassie Stain obtained from Generon

2.2 Buffers and media

Luria-Bertani (LB) media was prepared by dissolving the following pre-mixed components in deionised water: 0.5 % (w/v) yeast extract, 1 % (w/v) bactotryptone and 0.5 % (w/v) NaCl pH 7.4 (1 L of LB was used for protein purification, 5 ml for DNA purification and 100 ml for overnight cultures). All media were then autoclaved for 20 minutes at 121 °C and 138 kPa.

LB agar plates were prepared by adding 1.5 % (w/v) agar to LB medium before autoclaving. Colony selection was achieved by adding kanamycin to the cooled molten agar to a final concentration of 30 μg/ml, unless stated otherwise.

In all the unfolding, refolding and ligand binding experiments the same assay buffer was used, except for temperature-induced protein denaturation. The standard assay buffer was prepared by dissolving 0.05 % (w/v) DDM and 2 mM β-mercaptoethanol in 50 mM Tris buffer at pH 7.4. Stock concentrations of GuHCl and ligands were also prepared in this buffer. GuHCl stocks were adjusted to pH 7.4. The temperature denaturation buffer solution used was identical to the assay buffer with the exception of the Tris buffer, which was replaced by 50 mM HEPES solution.

2.3 Plasmids and *E.coli* strains

The GlpT gene used in this work was isolated and provided by Dr. Heather Findlay (King's College London) by subcloning the mentioned gene from the *E.coli* genome into the later used vector.

All the changes to the GlpT gene were done using plasmid pST-Blue-1 as a vector, later transformed into either commercially bought or grown in-house One Shot Top10 Chemically Competent *E.coli*. The GlpT gene was cloned into the plasmid pET-28a, modified with a 10-His tag, and transformed into either commercially bought or grown in house One Shot BL-21-AI Chemically Competent *E.coli* for protein overexpression. All plasmids used conferred kanamycin resistance for product selection and were supplied from Novagen. Both strains of *E.coli* were purchased from Invitrogen Life Technologies. The competent cells grown in-house were made according to the Promega protocol described in section 2.4.3.

2.4 Molecular biology

Restriction endonucleases and their relevant buffers were purchased from New England Biolabs. All oligonucleotides used for mutagenesis were synthesised by Eurofins MWG Operon using high purity salt-free purification, and delivered in a lyophilised form. Pfu polymerase was used for all PCRs, and was obtained from either Thermo Fisher Scientific or Promega. Gene ruler 1 kb DNA ladder and DNA loading buffer were supplied by the Thermo Fisher Scientific. Diagnostic and preparative DNA electrophoresis was carried out in a 1 % agarose (w/v) gel made in either purchased or self-made TBE (Tris/Borate/EDTA) buffer (45 mM Tris-Borate, 1mM EDTA). The DNA staining was accomplished by either addition of ethidium bromide or GelRed (Biotium, Inc.). DNA images were visualised using Amersham Imager 600 using 520 nm excitation fluorescence mode. DNA sequencing was carried out by Eurofins MWG Operon using the Value Read service. Concentration and quality of DNA purification was determined

using NanoVue Plus™ Spectrophotometer (GE Healthcare). Presence of RNA and protein impurities was estimated by recording OD_{260} /OD_{280} and OD_{260} /OD_{230} ratios, respectively.

The following commercially available kits were used for DNA purification, and used according to the manufacturer's instructions:

QIAprep Spin Miniprep Kit	Qiagen
QIAprep Spin Midiprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
GeneJET Plasmid Miniprep Kit	Thermo Ficher Scientific

2.4.1 QuickChange mutagenesis

All site-directed mutations with the exception of V43A, L268A and L44A were introduced using QuikChange PCR protocol in pST-Blue-1 plasmid using an Arktik™ Thermal Cycler (Thermo Fisher). PCR reactions were optimised by varying the amount of PCR template and the magnesium concentration. Primers and the individual PCR conditions can be found listed in the Appendix A. All PCRs contained the following reaction mixture: 1 µl dNTPs (final concentration of 0.2 mM), 5 µl of 10x Pfu buffer 1 µl of Pfu polymerase, 1.25 µl of both forward and reverse primer (final concentration 0.25 nmol/µl), 1 µl of DMSO (final concentration 2%), 1 µl of the GlpT gene template (final concentration 25-50 ng/µl), varied concentration of $MgSO_4$ (final concentration 1-4 mM), and molecular biology grade filtered water from Sigma Aldrich was added to volume of 50 µl. The PCR mixture was amplified in a thermocycler, starting with a 1 minute incubation at 95 °C to completely separate the double-stranded template. This was followed by 25-30 cycles of the following steps: 95 °C for 30 seconds, then by 30 s at the annealing temperature specific to the primers used, and then a final extension step at 12 minutes at 72 °C. For the I30A, I256A, L26A and L267A mutagenesis reactions a touch-down PCR was performed, where after every 5 cycles the annealing temperature was lowered by 2 °C,

starting from primer's theoretical melting temperature. At the end of all the PCR reaction programs, samples were subjected to a final 12 minute extension step. The unmethylated PCR product was then selected by a 2 hour incubation at 37 °C with 1 µl DpnI restriction enzyme, to remove the template from the reaction.

PCR product was analysed on the 1% agarose gel. The resulting gel presented a single band at 5.1 kb indicating amplification of the template with several less intense bands showing DpnI digestion fragments of the methylated template DNA. The PCR product of the right size was then excised and purified using QIAquick Gel Extraction Kit, and subsequently transformed into Top10 chemically competent *E.coli* using the usual transformation protocol.

2.4.2 'Round the horn mutagenesis

A variation of the standard mutagenesis protocol was used on less efficient mutation sites, where primer would bind non-specifically due to long repeats of a single nucleotide or a high predisposition to form dimers or secondary structure. Moore mutagenesis or 'Round the horn PCR, developed by S. Moore (https://openwetware.org/wiki/%27Round-the-horn_site-directed_mutagenesis), makes use of different primer design, shown in Appendix A. Due to the later ligation step, primers used were first phosphorylated for 1 hour at 37 °C in the following 50 µl reaction mixture: 5 µl 10X kinase reaction buffer, 1 µl 50 mM $MgSO_4$, 5 µl primer 10 µM final concentration), 1 µl of 100 mM ATP and 1 µl T4 polynucleotide kinase. The reaction was then heat deactivated at 95 °C for 5 minutes.

All 'Round the horn PCR reactions contained the following: 1 µl dNTPs (final concentration of 0.2 mM), 5 µl of 10x Pfu buffer 1 µl of Pfu polymerase, 1.25 µl of both phosphorylated forward and reverse primers, (final concentration 0.25 nmol/µl), 1 µl of DMSO (final concentration 2%), 1 µl of the GlpT gene template (final concentration 25-50 ng/µl); and varied concentration of $MgSO_4$ (final concentration 1-4 mM). The remaining volume was completed by the molecular biology grade filtered water from

Sigma Aldrich to the volume of 50 µl. PCR mixture was amplified in a thermocycler, starting with a 1 minute incubation at 95 °C to completely separate the double-stranded template, followed by 25-30 cycles of melting temperature of 95 °C for 30 seconds. This was followed by a varied annealing temperature, depending on the reaction done for 30 seconds and an extension time of 12 minutes at 72 °C. The PCR reaction was then subjected to the final 12 minutes extension step. The unmethylated PCR product was then selected by a 2 hour incubation at 37 °C with a 1 µl DpnI restriction enzyme.

PCR product was analysed on the 1% agarose gel. The resulting gel presented a single band at 5.1 kb indicating amplification of the template with several less intense bands showing DpnI digestion fragments of the methylated template DNA. The PCR product of the right size was then excised and purified using QIAquick Gel Extraction Kit and subsequently used in an overnight ligation reaction completed at 16 °C containing the following: 3 µl of 10 x T4 DNA Ligase buffer, 1 µl of T4 DNA Ligase, 100 ng of the PCR product. Deionised water was then added to the solution to make the volume up to 30 µl. Ligated product was the transformed into Top10 chemically competent *E.coli* using the usual transformation protocol, detailed in section 2.4.4.

2.4.3 Promega protocol for competent *E.coli*

A single colony from the agar plate or from glycerol stock was incubated in a 10 ml of the overnight culture. On the following day, 2.5 ml of the overnight culture was added to the 250 ml of the fresh LB and grown at 37 ºC until OD_{600} of approximately 0.5 was reached. Cells were then harvested for 5 minutes at 4500 g at 4 ºC. The pellet was then resuspended in a 100 ml of TFB1 buffer containing: 15% glycerol adjusted to pH 5 with acetic acid, 30 mM potassium acetate, 10 mM calcium chloride, 50 mM manganese chloride and 100 mM rubidium chloride. The resuspended pellet was then incubated on ice for 5 minutes. Cells were harvested again at 5 minutes at 4500 g at 4 ºC and then resuspended in a 10 ml of TFB2 buffer containing: 15% glycerol adjusted to pH 6.5 with potassium hydroxide, 10 M MOPS (3-(N-morpholino)propanesulfonic acid), 75 mM

calcium chloride, and 10 mM rubidium chloride. After 1 hour of incubation on ice, cells were aliquoted into 200 µl fractions and frozen in liquid nitrogen for storage at -80 ºC.

2.4.4 Transformations

Chemically competent *E.coli* were either purchased from Promega or grown according to the Promega protocol. Aliquots of competent *E.coli*, were defrosted on ice followed by the addition of 30-100 ng of plasmid DNA. The mix was incubated on ice for 20 to 30 minutes and then subjected to a heat shock at 42 °C for 45 seconds followed by a further incubation on ice for 2 minutes. After the transformation, 200 µl of fresh pre-warmed LB media was added, and transferred to the 37 °C incubator, with shaking, for 1 hour. The transformed *E.coli* was then spread onto LB agar plates containing 30 µg/ml kanamycin and incubated overnight at 37 ºC. To make glycerol stocks, the *E.coli* was re-seeded into fresh LB and grown until OD_{600}. The cells were then harvested at 4000 g for 10 minutes. The pellet was then resuspended in 10% glycerol and stored at – 80 ºC. After thawing, the glycerol stock transformed *E.coli* were grown on Kanamycin inoculated petri dish before being used for large scale protein expression, detailed in section 2.5.1.

2.4.5 Subcloning

All mutagenesis steps were performed in the pST-Blue-1 vector. Once a successful mutation had been confirmed by DNA sequencing the entire gene was subcloned into the expressing vector pET-28a using both XbaI and XhoI endonuclease restriction enzymes for 1 hour at 37°C. The 1.3 kb GlpT gene insert was isolated using DNA gel electrophoresis, and purified using the gel extraction kit according to the manufacturer's instructions. The empty 5.3 kb expression vector pET-28a, without an insert, was treated in the same manner as mentioned above. The purified insert was then ligated into the empty pET-28a vector at a 3:1 molar ratio respectively, using T4 DNA ligase in an overnight reaction at 16 ºC. The ligation reaction was transformed into *E.coli* Top10 cells using the standard heat shock protocol. The resulting *E.coli* colonies' plasmids were then

isolated and screened with an XbaI/XhoI digestion for the presence of the insert. The positive hits contained bands corresponding to a digested 5.3 kb vector and 1.3 kb size insert visualised on a 1% agarose DNA gel. The DNA was then sequenced to ensure the correct orientation of the insert. Successfully ligated plasmid was then transformed into BL-21-AI Chemically Competent *E.coli,* ready for protein expression.

2.4.6 Mutagenesis of single Trp mutants

Single Trp mutants of GlpT were made in two stages. Firstly, a tryptophan-free GlpT was designed by substituting all 15 trp residues with phenylalanine. The gene was synthesised by MWG Eurofins and delivered in a pEX-A2 vector with ampicillin resistance. The gene was then subcloned into pST-Blue-1 vector, where the rest of the mutagenesis steps were performed as described in the previous section. The individual single trp mutants were generated by substituting M107, F148 and V332 amino acids with Trp coding codon TGG in the tryptophan lacking gene variant of GlpT. The presence of the mutations was confirmed by sequencing before subcloning into the pET-23a vector for protein expression.

2.4.7 Mutagenesis of cysteine mutants

To further investigate the stability of the WT and alanine mutants, a cysteine mutant was designed to monitor the degree of unfolding using fluorescence recovery with rhodamine unquenching during the unfolding process. Out of 7 cysteines present in the GlpT gene, C215 is the only cysteine that is exposed to the protein exterior and could be reactive with the thiol-reactive dye. As a first step, the C215 was substituted with serine. C215S GlpT construct was then used to add cysteine residues, by substituting S61 and V283, present on cytoplasmic loops between the first helices of the first and second domain, respectively. The reaction was transformed into Top10 *E.coli* and then screened using PCR. The colony with a 1.3 kb PCR fragment was purified and sent for sequencing. The plasmid containing all 3 mutations was confirmed by sequencing and digested with XbaI

and XhoI restriction endonucleases, and the gene subcloned into pET-28a vector for protein expression.

2.4.8 DNA sequencing and primer synthesis

DNA sequencing was carried out by Eurofins MWG Operon using the Value Read service, based on cycle dideoxy chain termination sequencing technology. Sequences were analysed using Omega Clustal alignment algorithm and examination of a supplied chromatogram.

All oligonucleotides used for the mutagenesis reactions were synthesised by Eurofins MWG Operon using high purity salt-free purification method and delivered in lyophilised form. Delivered samples were solubilised in molecular biology grade filtered water from Sigma Aldrich to standard stock concentration of 100 nmol/µl and kept at -20°C.

2.5 Protein expression and purification

Throughout the protein purification protocol, all protein samples were kept on ice, and purified at 4°C in the in-house cold room. All protein buffers contained: 50 mM Tris buffer pH 7.4, 2 mM β-mercaptoethanol and 10 % (w/v) glycerol. Variation and detail of these buffers is further listed in the appropriate sections.

2.5.1 GlpT overexpression in *E.coli*

For a standard, large scale protein expression, 6 L of kanamycin inoculated media were used to grow GlpT. The use of baffled flasks increases aeration of media, contributing to the increased total bug growth and consequently increased protein yield per protein purification. No improvement in the protein yield was observed when the *E.coli* were grown in TB media, regardless of increased OD_{600} values measured throughout the bacteria growth.

A transformed *E.coli* colony containing GlpT in pET-28a from an LB agar plate, or a glycerol stock, was used to seed a 100 ml LB media inoculated with 30 µg/ml kanamycin, and incubated overnight at 37°C and 220 rpm. The overnight culture was then added at a 1:100 ratio to six 1 L flasks of pre-warmed LB media. The cultures were grown at 37°C with shaking at 220 rpm, and the growth was monitored at approximately 30 minute intervals, corresponding to the *E.coli* mean doubling time. The growth culture was induced with 1 mM IPTG and 0.1% arabinose at approximately 2.5 hours into the bacteria growth, when the OD_{590} reached 0.8 AU. The growth was continued for another 2 hours until it became stationary. At this point, the bacterial cells were harvested by centrifugation at 5000 g for 45 minutes at 4 °C (using Beckman Coulter J6-MI Centrifuge). The supernatant was discarded and the pellets were re-suspended in 200 ml PBS. The cells were then re-harvested at 5000 g for 10 minutes (Beckman Coulter Avanti J-26 XP). The supernatant was discarded and the bug pellet was weighted and re-suspended in 50 ml PBS with 10 mM β-mercaptoethanol and a protease inhibitor cocktail tablet (Roche Applied Science). The cells were frozen and stored at -20°C freezer until further use.

2.5.2 Protein purification

The frozen cells were defrosted and incubated with DNase I for 10 minutes at room temperature. The cells were then lysed using a cell disruptor (Constant Systems Ltd.) with an attached Thermoflex 1400 cooling unit (Thermo Ficher Neslab), followed by membrane sedimentation at 100,000 g for 30 minutes at 4°C (Beckman Coulter Optima XPN-90 Ultracentrifuge with type 70 Ti rotor). The membrane pellet was re-suspended in solubilisation buffer (50 mM Tris buffer pH 7.4, 10 mM β-mercaptoethanol, 200 mM NaCl, 20 mM imidazole, 10 % (w/v) glycerol, 2 % (w/v) DDM, 100 mM PMSF with the addition of EDTA free protease inhibitor cocktail tablet (Roche Applied Science)), and solubilised for 2 hours at 4 °C with stirring. 30 to 50 ml of solubilisation buffer (50 mM Tris buffer pH 7.4, 10 mM β-mercaptoethanol, 200 mM NaCl, 20 mM imidazole, 10 % (w/v) glycerol, 2 % (w/v) DDM, 100 mM PMSF with the addition of EDTA free protease

inhibitor cocktail tablet (Roche Applied Science)) was used, depending on the weight of the *E.coli* pellet at the end of the bacteria growth. solubilisation buffer (50 mM Tris buffer pH 7.4, 10 mM β-mercaptoethanol, 200 mM NaCl, 20 mM imidazole, 10 % (w/v) glycerol, 2 % (w/v) DDM, 100 mM PMSF with the addition of EDTA free protease inhibitor cocktail tablet (Roche Applied Science)). The solubilised membrane fraction containing the overexpressed protein was then isolated by centrifugation at 100,000 g for 30 minutes at 4 °C. The resulting supernatant was then injected into a 50 ml superloop attached to an ÄKTA pure chromatography system. The supernatant was passed through the 1 ml HisTrap HP Ni^{2+} metal ion affinity column (GE Healthcare) at a rate of 1 ml/min. The HisTrap column was pre-equilibrated with 10 ml of binding buffer (50 mM Tris buffer pH 7.4, 2 mM β-mercaptoethanol, 10 % (w/v) glycerol and 0.05 % (w/v) DDM). The column was then washed with 40 column volumes of wash buffer (50 mM Tris buffer pH 7.4, 75 mM imidazole, 2 mM β-mercaptoethanol, 10 % (w/v) glycerol and 0.05 % (w/v) DDM), to remove all nonspecific bound and unbound contaminating protein. Depending on the next purification step, the protein was then eluted into either a 2 ml loop on the ÄKTA pure, or eluted directly into an Amicon Ultra 50 kDa MWCO (Millipore) spin concentrator.

2.5.3 Desalting

The elution from the 1 ml HisTrap HP Ni^{2+} column was concentrated to a volume of 2 ml using an Amicon Ultra 50 kDa MWCO (Milipore) spin concentrator at 3500 rpm for approximately 20 minutes in a Megafuge 1.0R centrifuge (Heraeus Instruments). The concentrating was stopped at 5 minute intervals to stir the sample in order to avoid local crowding of the protein at the filter piece that could start protein aggregation. The concentrated protein sample was then loaded onto 5 ml HiTrap Desalting column (GE Healthcare) equilibrated with 10 ml of binding buffer (50 mM Tris buffer pH 7.4, 2 mM β-mercaptoethanol, 10 % (w/v) glycerol and 0.05 % (w/v) DDM) at of rate of 1 ml/min. The protein fractions were collected was and their concentration was measured using the UV absorbance at 280 nm using Cary 300 UV-visible spectrophotometer (Agilent Technologies), using the extinction coefficient calculated by ExPASy ProtParam tool

(146). The sum of the total protein yield was used to estimate the total yield for the protein per 1 L of *E.coli* culture grown. The fractions collected were then aliquoted, or if required concentrated again for storage or for further purification using gel filtration column. All protein samples were snap frozen using the liquid nitrogen before being stored at -80 °C.

2.5.4 Gel purification

Using gel filtration after purification with a HisTrap column yields a more purified product, compared to the HiTrap Desalting column, at the expense of slight dilution and protein loss. The majority of GlpT purifies as a monomer. However, following elution from a HisTrap column, GlpT oligomers can be observed by SDS-PAGE analysis. To remove these oligomers and ensure artefact-free measurements during the sensitive kinetic experiments, an additional purification step was utilized. The HiLoad Superdex 200 PG gel filtration column (GE Healthcare) was equilibrated with 120 ml of degassed desalting buffer (50 mM Tris buffer pH 7.4, 2mM β-mercaptoethanol, 10 % (w/v) glycerol and 0.05 % (w/v) DDM). The protein sample was loaded into AKTA Pure (GE Healthcare) *via* a 2 ml loop. If the protein was previously spin concentrated, the protein was filtered through a 0.45 µm Minisart syringe filter (Sartorius) prior to loading the sample onto the column. The column was then run at 1 ml/min for 2 hours with 1 ml fractions collected. The purified GlpT in a DDM micelle could be found exiting the column after approximately 70 ml into the run, corresponding to molecular weight of roughly 72 kDa, determined *via* calibration using gel filtration standards found in figure 2.1. The eluted fractions containing the protein were identified and their concentration established using a 280 nm UV absorbance (Cary 300 UV-visible spectrophotometer (Agilent Technologies)). A chromatogram of the representative purification recorded throughout entire purification procedure can be seen in Figure 2.2. The sum of total protein concentration was used to estimate the total yield for the protein per 1 L of *E.coli* culture grown. The fractions collected were then aliquoted or if required, concentrated again for storage or further

purification using gel filtration column. All protein samples were snap frozen using the liquid nitrogen before being stored at -80°C.

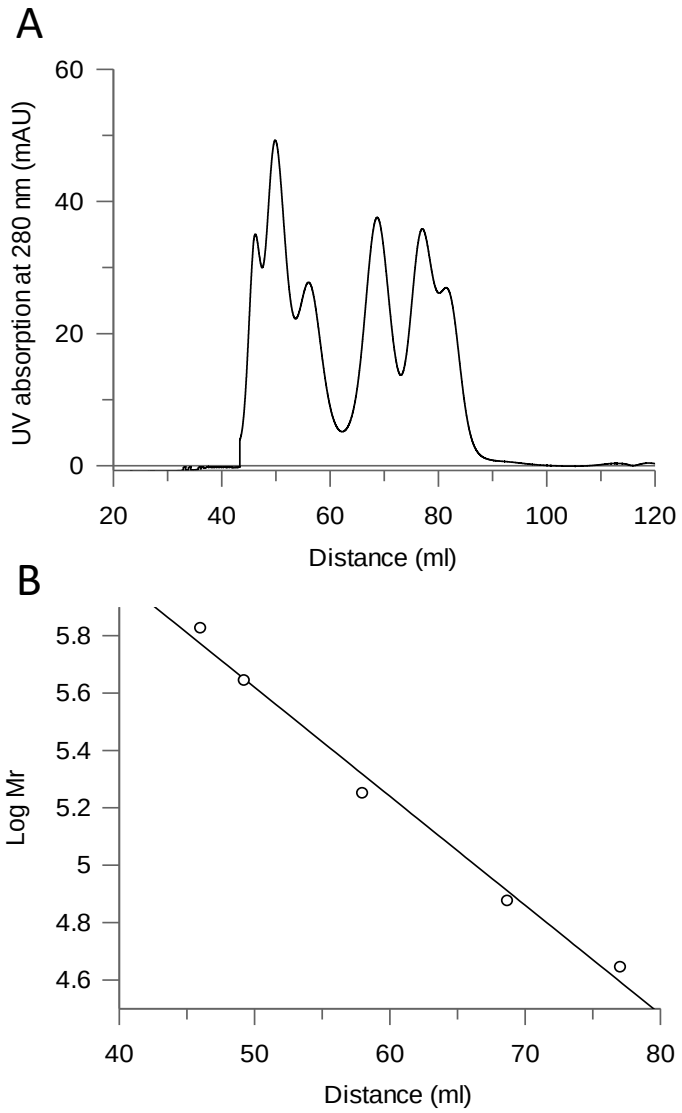


Figure 2.1 Gel filtration standards elution profile

The molecular size of GlpT in a DDM micelle was calculated using gel filtration column. Column calibration using molecular weight standards can be seen resolved in the elution chromatogram in panel A, the standards were supplied by GE Healthcare and contained thyroglobulin (669 000 Mr), ovalbumin (440 000 Mr), aldolase (158 000 Mr), conalbumin (75 000 Mr) and ferritin (44 000 Mr). The resulting data was collected and fit to a straight line equation, shown in panel B, with the point of y intercept equal to 7.5 ± 0.1 and gradient of -0.04 ± 0.01 Errors were taken from the standard error of the fit.

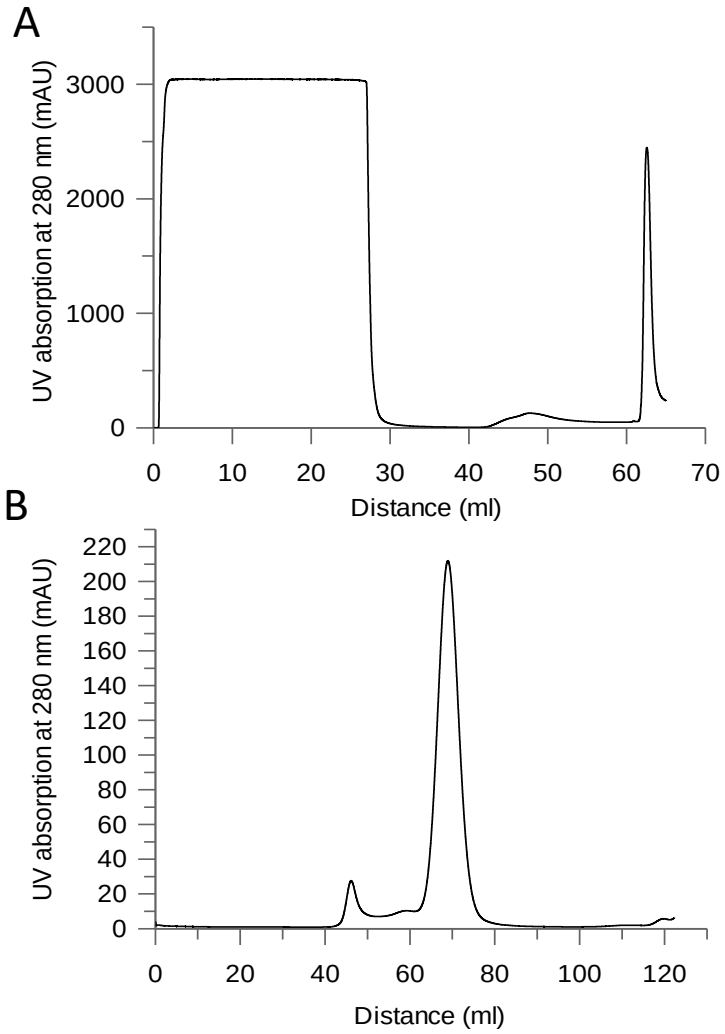


Figure 2.2 Chromatogram of GlpT WT purification

GlpT WT was purified from overexpressed *E.coli* culture with the solubilised cell lysate being purified *via* Ni^{2+} affinity column and gel filtration column. The first step of the purification involving Ni column is described in panel A, where in the first ml of purification procedure cell lysate is passed through the column, saturating the detector. After 30 ml of cell lysate the column was washed with moderate concentrations of imidazole. The sharp elution peak is collected in a 2 ml superloop, when the column is washed with 500 mM concentration of imidazole. The elute is loaded straight onto the gel filtration column and second purification step is initiated, shown in panel B. The sample is eluted as a monomer at 70 ml mark, with a small aggregated fraction at the void volume found at 40 ml. The fraction of the gel filtration peak representing the top concentration of the protein is then collected, concentrated and kept at -80 °C for further use.

2.6 SDS-Page analysis

Protein gel electrophoresis was used to determine the quality, efficiency and any potential differences in the purification protocol of the GlpT wild type and any of its mutants. The samples were run on a precast 10 % Tris-Glycine polyacrylamide gels (NuSep) containing: 100 mM Tris pH 8.25, 100 mM Tricine and 0.1 % SDS. Precision Plus Protein™ All Blue Standard protein marker (Bio-rad) was run with the samples to estimate the molecular size of the protein samples. The protein samples were mixed with 2x SDS loading buffer (50 mM Tris pH 6.8, 5 % β-mercaptoethanol, 12% glycerol, 3 % SDS and 0.01% bromophenol blue) in a 1:1 ratio prior to loading. Samples were then run at 190 V until the buffer ran to the bottom of the gel, for approximately 60 to 90 minutes. The purified GlpT band was observed just above the 37 kDa marker, as seen in the gel stained with the Coomassie dye in figure 2.3. The GlpT band does not appear at the predicted region of 52 kDa, which is due to the resistance of GlpT to SDS unfolding. The higher molecular weight bands represent dimer and trimer formations.

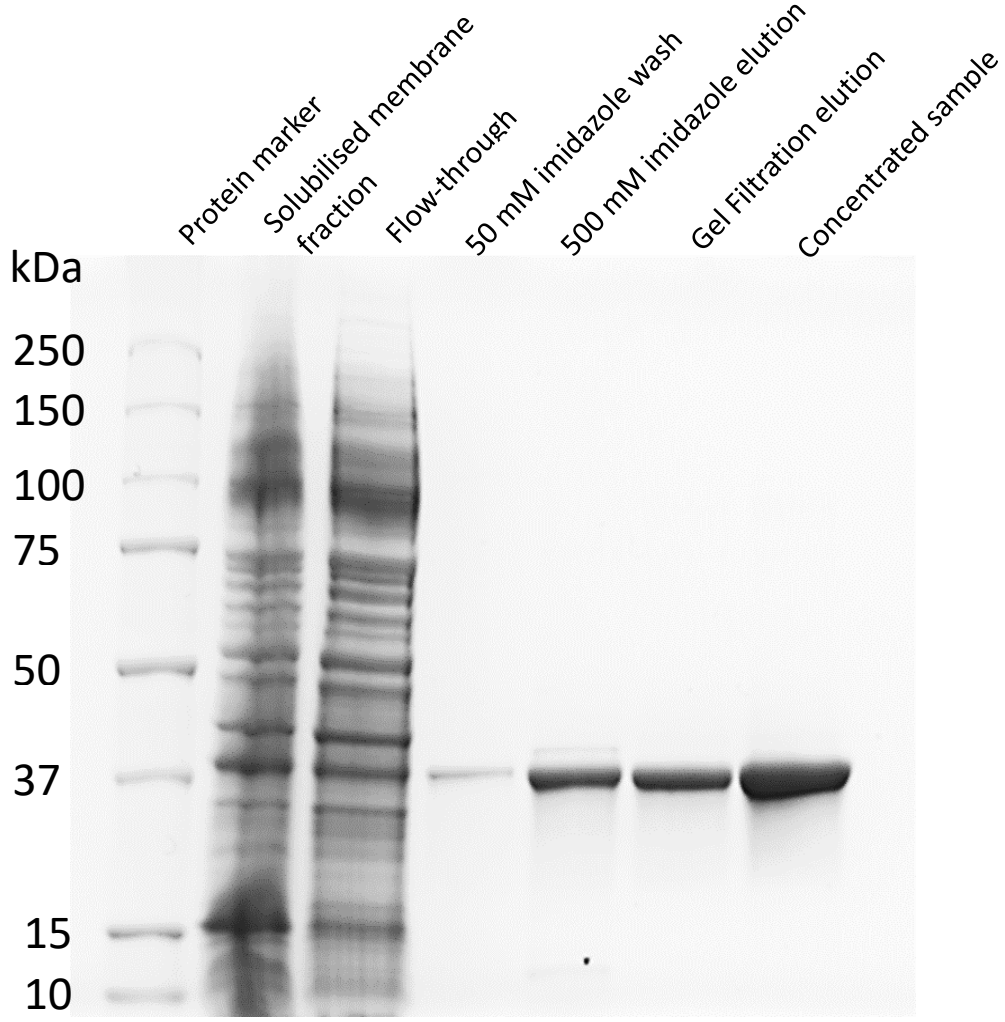


Figure 2.3 SDS-PAGE gel of GlpT WT purification visualised by coomassie blue staining

SDS-PAGE analysis was used to determine efficiency and purity of GlpT at various stages of purification. Samples of 10 µL were taken at each stage and run on gel as follows: All-blue protein standard (corresponding molecular weight of each band labelled on the left in kDa, solubilised membrane fraction derived from the cell lysate, flow-through of unbound fraction, 50 mM imidazole wash step, 500 mM imidazole elution of the Ni affinity column, top gel filtration fractions pooled together, spin concentrated fraction of protein before being stored at – 80 °C. Both Ni affinity (1 ml HisTrap HP Ni^{2+} metal ion affinity column) and gel filtration (HiLoad Superdex 200 PG gel filtration column) were carried out at a flow-rate of 1 ml\min, mounted on the ÄKTA pure chromatography system.

2.6.1 Coomassie staining

SDS-PAGE gels were visualised with Coomassie stain. The Quick Coomassie (Generon) was used to stain gels for 60 minutes. Background staining was eliminated by leaving the stained gel overnight in a deionised water. For gels with high background stain, gels were microwaved at full power for 10 seconds with frequent water changes. Stained gels were then imaged in Amersham Imager 600 using the trans-illumination picture.

2.6.2 Western blotting

SDS-PAGE gels were also visualised by western blot. The protein was transferred onto a Hybond-EXL nitrocellulose Protran membrane with 0.45 µm pore size (Amersham) at 45 A for 60 minutes, using a TE 70 PWR Semi-Dry Transfer Unit (GE Healthcare). The membrane and stacks of filter paper were soaked with transfer buffer (25 mM Tris base pH 8.3, 150 mM glycine and 10% (w/v) methanol) prior to transfer. To eliminate non-specific binding of the antibody, the membrane was blocked in 5 % milk and 0.05 % (w/v) PBS-Tween for 60 minutes at room temperature with rocking. Horseradish peroxidase conjugated monoclonal anti-polyHis antibody from mouse was then added at a 1:2000 dilution to 5 % milk in 0.05 % (w/v) PBS-Tween solution for further 60 minutes at room temperature with rocking. The unbound antibody was then washed off with 3 1 minute washes of 0.05 % (w/v) PBS-Tween followed by 3 10 minute washes of 0.05 % (w/v) PBS-Tween. The nitrocellulose membrane was then incubated for 1 minute at room temperature with LumiGLO (Cell Signalling Technology), in a total volume 10 ml. The nitrocellulose membrane was then dried and visualised in Amersham Imager 600 using cheminolumiescence settings with automatic exposure setting unless stated otherwise.

2.7 Fluorescence spectra

All fluorescence spectra were measured in a spectrofluorometer Fluoromax-4 (Jobin Yvon, ISA Instruments, S.A Inc). Each sample was excited at both 280 nm and 295 nm, and their emission recorded at 0.25 nm increments between 300 and 450 nm with an

integration time of 0.1 s. Both excitation and emission slits were set to a 5 nm bandwidth. A 4 mm pathlength Starna Quartz cuvette was used, with the temperature set to 25 °C with a temperature tolerance of 0.1 °C, unless stated otherwise. The baseline of the buffer and denaturant were subtracted from each spectrum before the analysis of the scan, as both urea and guanidinium chloride have an emission band at 330 nm. All of the fluorescence data analysis was done in GraFit software (Version 5, Erithacus Software Limited). For an accurate estimation of fluorescence intensity at the wavelength maxima and position of the wavelength maxima, a normal distribution curve was fitted to the data of the obtained spectra. If the fitting of the curve was acceptable, the position of the peak was used to estimate the wavelength maxima, as well as maximum fluorescence. The normal distribution fit equation can be seen below, with μ being mean and **σ** symbol used for standard deviation.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

2.8 Circular dichroism spectra

All circular dichroism measurements were taken on an Aviv Circular Dichroism Spectrophotometer, Model 410 (Biomedical Inc, Lakewood, NJ USA), with a custom telescopic detector installed in order to reduce possible scattering artefacts. CD spectra were recorded at 1 nm increments with 3 repeated scans between 270 and 185 nm and integration time of 1 s, unless stated otherwise. Measurements containing urea were taken between 270 and 200 nm due the absorbance of urea in the lower wavelength region. The type of the cell, as well as its pathlength depended greatly on the demands of the experiments and the protein concentration used. The cell pathlength varied between 0.02 and 2 mm. The details of the experiments are discussed where relevant.

2.9 Circular dichorism spectra analysis

All CD spectra were processed using the CDtool program (version 1.4) (142). The buffer baseline of the matching cell pathlength was subtracted from each spectrum. The spectra were also zeroed to a common set baseline between 253 and 260 nm. The raw spectra were also automatically smoothed by an inbuilt Savitsky-Golay filter. The buffer baseline of the matching cell type and cell pathlength was subtracted accordingly from each spectra to aid in more accurate deconvolution of the data.

To analyse multiple experiments and to allow comparison, the data was converted to a standardised unit of mean residue ellipticity expressed by $deg.cm^2.dmol^{-1}$ units. Expression used to convert raw milidegrees signal (mdeg) into the mean residual ellipticity in $deg.cm^2.dmol^{-1}$ (MRE), is shown below, with mean residue weight (MRW) in kDa, pathlength, L, in cm and concentration, c in $mg.ml^{-1}$, the MRW, mean residual weight is protein dependant and in GlpT is equal to 110 kDa.

$$MRE = \frac{mdeg\ x\ MRW}{10\ x\ c\ x\ L}$$

Alpha-helical proteins have two negative peaks at 222 nm and 208 nm. The change of peak at 222 nm rather than 208 nm was used as indication of the structure change. The 208 nm peak of GlpT is much shallower than 222 nm peak, especially at the higher urea concentrations. Furthermore, the increasing urea concentrations interfere with measurements recorded below 200 nm, due to the urea's own spectroscopic properties.

2.10 Circular dichroism deconvolution

Insight into the folded structure of proteins can be estimated *via* comparison of acquired CD spectra to a large database of known crystal structures and their respective CD spectra. The dichroweb web server (http://dichroweb.cryst.bbk.ac.uk/, 147) was used to estimate the secondary structure of GlpT (148). The process of spectra deconvolution

requires a wide wavelength range, inaccessible for denatured protein and limited by low protein concentration and scattering effects. The membrane protein specific SMP180 dataset 4 was used for samples with spectra recorded to 180 nm (149). Samples with spectra recorded to only 190 nm were analysed with a less accurate estimations from dataset SP175 (150). The fitting was considered good when the NRMSD value did not exceed 0.1.

2.11 Circular dichroism equilibrium unfolding

A range of concentrations of urea were made from a 10 M stock solution. The stock was made by dissolving urea in a premade buffer (50 mM Tris pH 7.4 and 0.05% DDM). Aliquots of GlpT were defrosted on ice, and added to the denaturant to make up the final denaturant concentration in a total volume of 50 µl. Each unfolding reaction was incubated for 5 minutes at 25 °C. The total protein concentration was kept at a range between 0.3 and 0.5 $mg.ml^{-1}$, giving a good quality CD spectra. Both rectangular and circular Suprasil demountable cells were used in acquiring unfolding spectra. Depending on the quality and protein concentration pathlengths of 0.1 mm or 0.2 mm were used. The sample was equilibrated for 1 minute in the sample holder.

2.12 Fluorescence equilibrium unfolding

The protein sample was incubated with increasingly higher concentrations of denaturant, either urea or GuHCl, in a total sample volume of 300 µl, for 5 minutes at 25 °C. The final protein concentration was kept constant throughout the experiments, at 0.025 $mg.ml^{-1}$. The sample was equilibrated for 1 minute in the sample holder prior to measurement.

2.13 Circular dichroism equilibrium refolding

The method described in section 2.12 was also used to unfold GlpT prior to refolding. The protein was denatured in 8M urea for 5 minutes at 25 °C, followed by a rapid ten-fold dilution with refolding buffer (50 mM Tris, 0.05% DDM). The urea concentration

varied between the samples in order to give a range of final urea concentrations after the rapid dilution, ranging from 0.8 to 8 M. The final protein concentration varied between 0.01 $mg.ml^{-1}$ to 0.04 $mg.ml^{-1}$. The sample was then incubated for 10 minutes at 25°C. To acquire a good CD signal the spectra of the refolded samples were measured in a 0.5 mm quartz rectangular or round Suprasil demountable cells (Hellma Analytics).

2.14 Fluorescence equilibrium refolding

To achieve a final protein concentration of 0.025 $mg.ml^{-1}$ in a sample volume of 400 µl, the starting 1 $mg.ml^{-1}$ of protein was unfolded in 8 M urea for 5 minutes at 25 °C. Afterward the unfolded mixture was diluted into refolding buffer with decreasing concentrations of urea, to give a final range of urea between 8 M to 0.8 M. Each emission spectra was subtracted against the background emission of each of urea concentrations using the same volumes without the protein present.

2.15 Circular dichorism thermal denaturation

Thermal denaturation was measured by heating the protein from a temperature of 5 °C to 95 °C degrees in 2 °C steps. The detector was set at 222 nm to monitor the loss of alpha helical structure. The sample was allowed to equilibrate for 30 seconds at each temperature step before the scan was taken. Instead of 50 mM Tris pH 7.4, 50 mM HEPES buffer was used, due to the temperature pH sensitivity of Tris buffer. The protein sample was measured at a concentration of 0.1 $mg.ml^{-1}$ in a rectangular 1 mm pathlength cell. The high protein concentration and longer pathlength used ensure a good CD signal during the large amount of structure loss at temperatures above.

2.16 Crystal structure images

All figures featuring crystal structures were rendered using The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC. The crystal structures were acquired from the protein data bank using PDB codes stated below: 1pw4 (GlpT, Uniprot: P08194,

97). UCSF Chimera (http://www.rbvi.ucsf.edu/chimera) was used to model the missing interdomain loop. The model with the lowest energy was used in making GlpT molecular images in PyMOL.

2.17 Two-state analysis

The parameters from the spectroscopic measurements such as wavelength of the fluorescence maximum, the fluorescence emission and the mean residual ellipticity recorded at 222 nm were used to monitor the extent of the folded structure at different denaturant concentrations. These results were then fitted into two-state model to find the free energy of unfolding in the absence of denaturant (ΔG_U^{H2O}) using the equation

$$y = start + range \frac{exp\,(-m(Cm - x)/-RT)}{1 + exp\,(-m/(Cm - x)/-RT)}$$

derived from: $\Delta G_U^{H2O} = Cm \times m$ and $\Delta G = -RT \ln Keq$, where $K_{eq} = \frac{f_{unfolded}}{f_{folded}} = \frac{f_{unfolded}}{1 - f_{unfolded}}$

The zero denaturant containing point, where protein is thought to be folded, was chosen as a start parameter, while the range is described as an amount of total denaturation between the points of zero denaturant and maximum denaturant concentration x, The R is the gas constant (1.987 x 10^{-3} kcal/K.mol), T is the absolute temperature, x describes the denaturant concentration with m being the gradient of the transition slope

2.18 Markwell-Lowry

A modified version of the Lowry quantification assay (151) designed for the membrane and lipoprotein preparations was used in estimating protein concentration for the single tryptophan assays, using bovine serum albumin as a standard. Protein samples of unknown concentration and the standard were pipetted into a 0, 5, 10, 15, 20, 25, 50 µL

series of volumes. The samples were then fully solubilised by addition of 50 µL of 10 $mg.ml^{-1}$ sodium deoxycholate. After 1 minute incubation at room temperature, 1 ml of 10 % (w/v) trichloroacetic acid was added to the samples. The precipitate was spun down at 10 000 g for 5 minutes, and the supernatant discarded. Protein precipitate was solubilised with 1 ml alkaline copper reagent. The reagent was made of solution A (200 mM anhydrous Na_2CO_3,100 mM NaOH, 7 mM potassium sodium tartrate, 1% (w/v) SDS), and solution B (4% (w/v) $CuSO_4.5H_2O$) premixed in a 100:1 ratio, respectively. Afterwards, 10 µL of the 1 M Folin and Ciocalteu's phenol reagent, was added to the solubilised mixture and stirred immediately. The sample was left to incubate at room temperature for 1 hour before the UV absorbance readings were taken at 750 nm in 1 ml quartz cell. Protein amount was then determined according to the known bovine serum albumin concentrations by interpolation.

2.19 Rhodamine conjugation

Tetramethylrhodamine-5-maleimide (TMR), purchased from Thermo Fisher Scientific, was used in labelling of the double cysteine mutant of GlpT. The 0.5 M stocks of TMR were kept frozen at -20 °C in DMSO. Prior to conjugation, the protein was diluted to a concentration of 20 µM in buffer containing 50 mM Tris at pH 7.4, 200 µM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and 0.05 % (w/v) DDM in a total reaction volume of 250 µL. The protein was incubated in TCEP on ice for 10 minutes to reduce any disulphide bonds prior to dye conjugation. The protein sample was then transferred to a darkened light-safe 0.5 ml eppendorf along with the excess of the TMR solution. It was found that an overnight incubation at 4°C with the 40-fold molar excess of the TMR label provided the best labelling efficiency of 1.7 moles of labelled dye to moles of protein.

2.19.1 Free dye removal

The fluorescent dye removal resin and columns (Pierce dye removal columns) were acquired from Thermo Fisher Scientific. All procedures were performed at room

temperature. First, 400 µL of the uniform suspended resin was added to the spin column and centrifuged at 1000 g for 45 seconds to remove the resin storage solution. Next, 400 µL of 50 mM Tris at pH 7.4, 200 µM TCEP and 0.05 % (w/v) DDM were added to equilibrate the column before another 1000 g spin for 45 seconds. Finally, 250 µL of the labelling reaction were pipetted into the resin and left to incubate for 1 minute. The column was then spun down for further 45 seconds at 1000 g. The eluted sample was collected and the process was repeated again with fresh column and resin to remove any free TMR label that might have been left. The purified labelled protein was protected from light and either immediately used or stored at -80 °C avoiding multiple freeze-thaw cycles.

2.19.2 Determination of degree of cysteine labelling

A Cary 300 UV-visible spectrophotometer (Agilent Technologies) was used to record the UV absorbance of the labelled samples. The sample was diluted in 400 µL of 50 mM Tris pH 7.4 and 0.05 % (w/v) DDM, and the UV absorbance at 280 and 541 nm was measured in a 1.4 ml UV-vis ES quartz cell (Aireka Cells).

In order to assess successful dye removal and efficiency of dye conjugation, the following formula was used:

$$\frac{A_x}{\varepsilon} \times \frac{MW\ of\ protein}{mg\ protein/mL} = \frac{moles\ of\ dye}{moles\ of\ protein}$$

A_x is the UV absorbance of the dye at the maximum wavelength, and ε is the extinction coefficient of the fluorescent dye used. Both the equation and the TMR 541 nm UV absorption maxima and 97 000 $M^{-1}cm^{-1}$ extinction coefficient were taken from the Invitrogen Molecular Probes™ website (www.thermofisher.com/molecular-probes-the-handbook).

2.20 Cysteine mutant fluorescence unfolding assays

The cysteine labelled sample was incubated at 25 °C for 5 minutes with increasingly higher concentrations of urea denaturant, ranging between 0 to 8 M, in a total sample volume of 300 µl. All reactions were performed in light-safe 1.5 ml eppendorf tubes. The final protein concentration used was kept constant throughout the experiments at 0.015 $mg.ml^{-1}$. The sample was equilibrated for 1 minute in the sample holder prior to measurement in a 4 mm pathlength Starna Quartz cuvette. The fluorescence emission spectra was recorded at 0.25 nm increments between 555 and 650 nm with an integration time of 0.1 second with the excitation wavelength set at 550 nm. Both excitation and emission slits were set to a 3 nm bandwidth. The difference in fluorescence intensity, as well as shift of the fluorescence maxima were estimated from the recorded fluorescence spectra as described in section 2.7.

2.21 Cysteine mutant circular dichroism assays

All CD spectra of cysteine mutant were taken as described before, with the exception of extra addition of 10 mM TCEP, to stop unwanted cross-linking of two cysteines. CD spectra of assay buffer containing TCEP was used to subtract each spectra. The decay of the 222 nm peak with the increasing concentrations of urea was collected and analysed as described in section 2.9.

.

Chapter 3

Unfolding of GlpT WT

3.1 Introduction

3.1.1 Unfolding of detergent bound α-helical proteins

Protein stability can be assessed by monitoring the reduction in protein's structure during denaturation, followed by estimation of free energy of unfolding. The free energy of the unfolding can be summarised as the difference between the free energies of two states, the folded protein and the denatured structure. A relatively simple method of stability assessment is *via* denaturation curve performed at steady state conditions (152). Measurements of unfolding at steady state give a relatively straightforward insight into the equilibrium thermodynamics of the unfolding pathway leading to an increased understanding of underlying origin of protein stability.

Protein denaturation can be brought about by either a change in temperature or addition of chemical denaturants. Thermal denaturation occurs when increasing temperature disrupts amino acid contacts and as a consequence disturbs the protein structure. Alternatively, chemical denaturation methods can be used with chaotropic agents as denaturants (153). Examples of useful protein denaturants are SDS, urea and GuHCl. Interactions between water and urea or GuHCl, have been found to be less favourable than with protein's amino acids (154). Interestingly, both denaturants interact with polar and non-polar surfaces of the protein. However, GuHCl was found to interact more favourably with polar residues than non-polar residues. Urea was found to act in an opposite manner, and interact more favourably with non-polar residues over polar residues (155). Still, the exact mechanism through which these denaturant molecules unfold protein structure is not well understood.

The unfolded state of the proteins cannot be described using single structure. Instead, the unfolded membrane protein may exist as a collection of different states with a range of varying micelle-protein structure complexes and different stabilities (156). Therefore, specific information on the structure of the protein is difficult to obtain and analyse. In this chapter, the degree of GlpT unfolding by both chemical and temperature methods is monitored by fluorescence and CD spectroscopy. The GlpT's reversibility ability is also assayed by measuring the return to the same structure content as before the denaturation.

The detergent systems do not fully reproduce the conditions of membrane proteins in their native cell environment. When denatured, many membrane proteins tend to form irreversible aggregates of multiple different conformations. Although, membrane proteins are considered to be unstable and ''frail'' in contrast to soluble proteins the membrane proteins resist the denaturation resulting in their unfolded state displaying elements of secondary structure (14). When considering the use of different denaturants, it is good to consider that the denaturant does not destabilise the lipid or detergent environment present in an experimental setup. The suitability of different denaturants is measured and discussed later in this chapter.

3.1.2 Two-state folding and free energy

In order, to apply thermodynamics equations that yield the free energy linked to protein folding, the protein needs to be shown to exist in a reversible equilibrium. Experimentally, the easiest circumstance is when protein folding occurs in a single step without any intermediates. In such two state reaction, only the unfolded and folded states exist without any intermediate states in-between. The reaction can then be represented as:

$$\text{Folded} \rightleftharpoons \text{Unfolded}$$

During denaturation experiments, higher concentrations of the denaturant drive this equilibrium towards the unfolded state. The relative populations of the folded and unfolded states describe the position of equilibrium and are defined by the equilibrium constant K_{eq}:

$$K_{\text{eq}} = \frac{f_{\text{unfolded}}}{f_{\text{folded}}} = \frac{f_{\text{unfolded}}}{1 - f_{\text{unfolded}}}$$

The likelihood that a reaction is populated by one state over another is described by free energy, ΔG, and is calculated by:

$$\Delta G = -RT \ln K_{eq}$$

Where R is the gas constant (1.987 x 10^{-3} kcal/K.mol) and T is the absolute temperature. Both fluorescence and CD measurements can be used to track the progress of the unfolding in the denaturation experiments. To find the free energy of unfolding in the absence of denaturant (ΔG_U^{H2O}) the following equation is used:

$$\Delta G = \Delta G_U^{H2O} - m[\text{denaturant}]$$

This model assumes linear dependence between the concentration of the denaturant and the free energy change, ΔG, where m defines an extent of the dependence of ΔG on the concentration of the denaturant (157). The denaturant concentration at the midpoint of the unfolding curve characterizes a state of equilibrium where ΔG=0. In that state it is possible to describe ΔG_U^{H2O} by following equation:

$$\Delta G_U^{H2O} = Cm \times m$$

At the midpoint of the reaction the population of the protein is split equally between the folded and the unfolded state, described by the Cm value. In the case of the simple two

state reaction described by a sigmoidal plot, the regions before and after the slope on the sigmoidal curve represent folded and unfolded states, where as the gradient of the transition slope of the denaturation curve describes the value m, and midpoint of the curve defines the value Cm (155). Alternatively, ΔG_U^{H2O} can be estimated by an extrapolation of a linear plot of free energy change ΔG at every denaturant concentration in the transition state against the denaturant concentration used.

The reversible equilibrium two-state system also implies that the free energy of unfolding is the same regardless of measuring it through the unfolding or refolding experiments, since both the forward and reverse reactions follow the same pathway. It is however more accurate to use the free energy calculated from unfolding rather than folding due to unfolding experiments being made out of fewer steps and the measurements are taken at higher and more accurate protein concentrations.

3.1.3 Reversibility of the unfolded state

The process of protein refolding from the denatured state would usually involve reduction or elimination of the denaturant that has led to protein unfolding. For example, by decreasing the temperature or dilution of the chemical denaturant, the protein is allowed to return to its native conformation, which was present before the process of unfolding. Generally, the temperature denaturation of large soluble proteins and membrane proteins results in a formation of irreversible unfolded aggregates. However, the reversibility of a chemically denatured state is frequently observed and has become a common method used in studying stability of membrane proteins. The most straightforward way of eliminating the chemical denaturant is *via* dilution or dialysis. When the concentration of denaturant decreases below that of a denaturing concentration, the protein recovers its structure.

Upon reduction of the denaturant concentration, the structure of the protein can be recovered. Although, even in the cases of full structure of recovery, it is possible that the

protein refolding has occurred *via* different pathway than the unfolding reaction (158). A possible cause for different pathway could be due to a change of renaturing environment, the state of a solubilising agent could have been affected by the denaturant or the refolding process itself. When the reaction of unfolding and refolding proceed through different reaction steps, it is no longer possible to interpret any thermodynamic equilibrium results, as the unfolding cannot be described as reversible, since the two pathways are not in equilibrium, and therefore the resulting free energy cannot be calculated. The diagram of different folding-reaction pathways presenting different outcomes is shown in figure 3.1. In this chapter, the extent of GlpT structure recovery upon the dilution of the denaturant is assayed and then compared to that of the other members of the MFS.

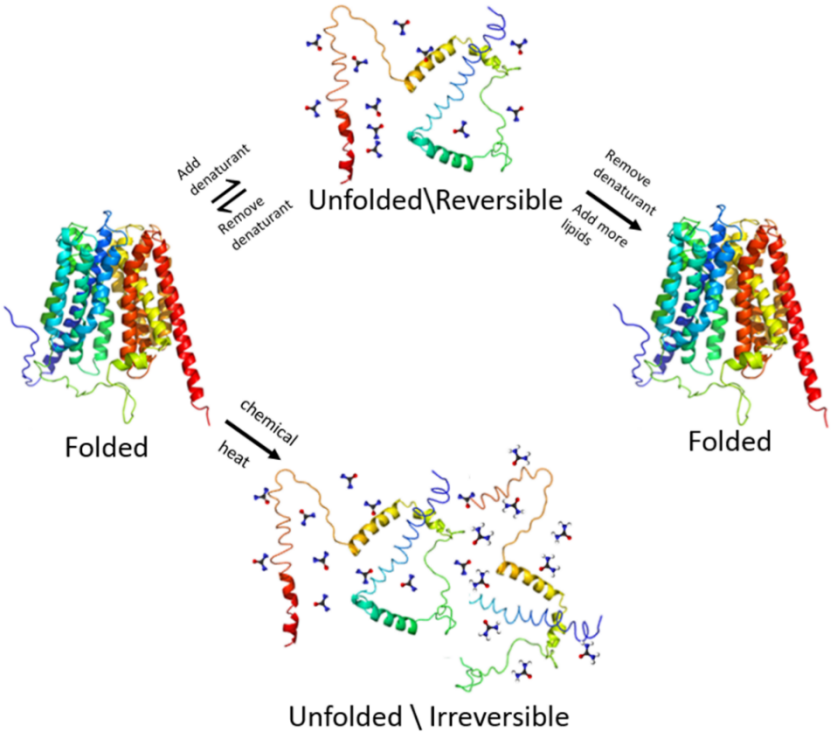


Figure 3.1 The differences between reversible and irreversible folding

The protein starts in the folded state, upon addition of denaturant and conditions it can assume different unfolded states. If the native structure of the protein returns when the denaturant is removed, then the process of unfolding is reversible. Frequently, the process of refolding is accompanied by addition of different renaturing lipids, in which case, the reaction is not reversible as it follows a different pathway even though the original structure is recovered. Diagram adapted from (159). Structures were rendered in PyMOL software.

3.1.4 Return of the protein function in the refolded state

The reversibility assay and the return of native protein structure is further examined by the additional test of measuring the return of protein function. Comparing the efficiency of substrate binding or substrate transport before and after denaturation adds an extra argument that the protein has regained some aspects of the structure that were present before the denaturation, since the substrate binding affinity is dependent on the conformation of its binding site. The ideal solution would be to test the ability of the transporter to shuttle substrates across the membrane into the liposomes. However, due to reconstitution inefficiency and addition of many extra factors, such as changing the solubilisation system during the reconstitution, the results collected from functionality assay may be inaccurate.

An alternative assay that measures affinity to substrate binding is also complicated. The transporters tend to have a low affinity to their substrates and the process of protein solubilising during purification may result in the binding pocket being obstructed by the detergent. The development of a GlpT functionality assay, based on measuring the affinity to substrate binding, is discussed later in this chapter.

3.1.5 The case for fluorescent labels

The process of unfolding of alanine mutants, described in chapter 4, is followed by changes of GlpT's structure, measured by CD. However, an additional method of following the steps of denaturation would expand the structural information about the denatured state. Measurements of protein fluorescence are one of the most commonly used experimental approaches in examining the transition phases present in the protein folding. This is due to the nature of fluorescence being sensitive to the changes in the fluorophore's environment. Moreover, fluorescence requires little material to produce a high signal-to-noise ratio. Compared to the circular dichroism recordings, which measures the global extent of the secondary structure, the fluorescence signal originates

from individual fluorophores and is directly linked to the local conditions surrounding the fluorophores.

Measurements of the intrinsic fluorescence of GlpT negate the benefit of linking fluorescence to the local environment due to well-spread presence of 15 tryptophan residues. This abundance is not uncommon in membrane proteins, as tryptophans were shown to stabilise the protein *via* interactions with lipid head groups (160). The following sections, are trying to lay down design principles of fluorescence reporters in GlpT.

3.1.6 Single tryptophan introduction

One method to determine the extent of unfolding in specific regions of the protein is through the substitution of single tryptophan residues into the tryptophan-free protein of interest (161). For example, the unfolding of a single tryptophan containing protein SNase can be followed by multiple observations of emission energy, anisotropy and fluorescence intensity (162, 163). In a more recent work (125), single tryptophan mutations were introduced to a tryptophan free template of LacY. The study identified regions of the protein that show a greater degree of unfolding, when compared to other single tryptophan sites in LacY. Although, the single tryptophan produces a fluorescence signal informing on the state of the local environment, for large proteins with multiple domains the fluorescent residue may be too far to report on the state of those regions. In that case, multiple single tryptophan substitutions need to be made in order to provide a complete picture of the protein unfolding reaction (164). Most membrane proteins contain more than one tryptophan residue. In the case of multi-tryptophan proteins a template needs to be made where the number of fluorescent reporters is reduced (165).

In order to use tryptophans as reporters of GlpT's unfolding, all 15 native tryptophan residues were replaced with phenylalanines. The fluorescence of phenylalanines (extinction coefficient around 200 $M^{-1}cm^{-1}$ at 280 nm) is much smaller than that of tryptophan (extinction coefficient around 5500 $M^{-1}cm^{-1}$ at 280 nm). At the same time

phenylalanine is the most conservative of mutations for replacing aromatic tryptophan. See figure 3.2 for the locations of tryptophan residues in WT GlpT as well as placement of single tryptophans in the mutant GlpT. None of the replaced tryptophans were reported in the previous GlpT studies as essential for protein function. Secondly, single tryptophan substitutions were made onto a tryptophan-free template, referred to in the text as Trp-free, with the intention that the fluorescence recorded from the protein would reflect the environment from the single fluorescence reporter. Allowing the comparison between different regions changes upon unfolding.

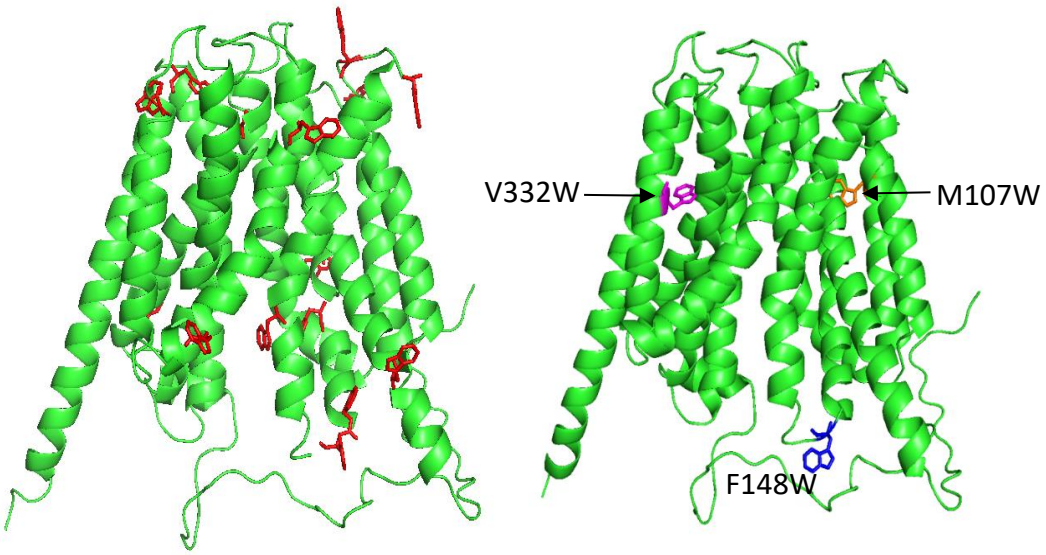


Figure 3.2 Representation of tryptophans in GlpT WT and single tryptophan variants

The structure on the left shows placement of all 15 tryptophans, in red, in GlpT WT. They are arranged in two planes, where they likely interact with lipid head groups in the native environment. On the right, positions of single-tryptophans are shown, in orange M107W, in blue F148W and finally in magenta V332W. The single tryptophans were placed to probe different environments, M107W is placed at the interhelical space in the first domain, F148 at the cytoplasmic interface and V332 in the interhelical space of the second domain. Structures were rendered in PyMOL software.

3.1.7 Rhodamine label

An alternative technique to single tryptophan constructs uses the interactions between two closely placed rhodamine labels to report on the change in the protein structure. When the two rhodamines are placed in a close proximity to each other, they form stacks causing fluorescence quenching between them *via* the mechanism of excitation coupling (166). During the conformational change, which occurs during protein unfolding, the distance between the two labels would change, therefore altering their fluorescence intensity. The majority of the studies that used double rhodamine labelled proteins were used to report major conformational changes in large soluble protein complexes (164, 168-169). An example includes a phosphate biosensor constructed using two rhodamine labels, which experiences a 20 fold increase in fluorescence upon phosphate binding (169). A similar method was used to measure the association of subunits during the dimer formation of *E.coli* ribosomal protein (170).

The distance between two labels is a major consideration in the design of this system, the labels are put in the most dynamic positions, so that during protein folding or unfolding the distance change between them would produce biggest effect on the fluorescence change. See figure 3.3 for representation of proposed sites of rhodamine labels. The accessibility of residues that would be labelled also needs to be considered. The thiol groups of cysteines need to be exposed for the thiol-reactive label to interact with them. The presence of solubilising agents may hinder labelling, if the molecule has a strong affinity towards a site that is close to the chosen cysteines.

In this work, two cysteine residues on the periplasmic side, S61C and V383C, were placed on the cytoplasmic loops connecting helices 1-2 and 7-8 in close proximity of 15.8 Å. An alternative placement of two labelling sites is present on the cytoplasmic side of residues S150 and A376, this construct was not tested due to time constraints. The stability of the S61C-V383C construct was tested and the fluorescence changes of rhodamine labels upon denaturation inspected.

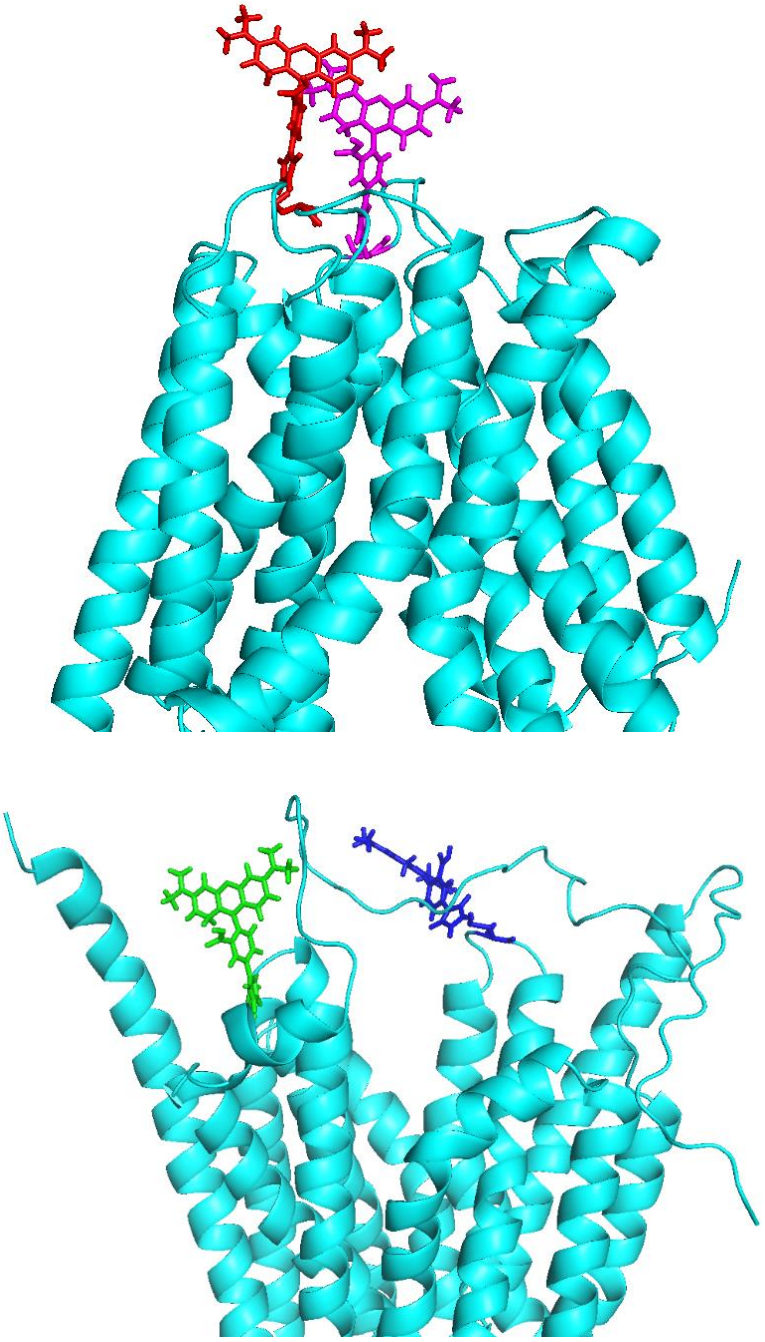


Figure 3.3 Representation of cysteine mutations in GlpT WT with the rhodamine labels

The structure on the top shows placement of rhodamine labels present on the endoplasmic face, on the bottom the alternative position of labels present on the cytoplasmic interface. The closest distance between α carbons of cysteine residues on the endoplasmic face (top figure) approximately equal to 15 Å, respective to the cytoplasmic face (bottom figure) 25 Å. Structures were rendered in PyMOL software.

3.1.8 Aims of the chapter

The aim of this chapter is to characterise both the unfolding and refolding extent of the GlpT WT using different denaturants. Different methods, such as, circular dichroism and fluorescence techniques are used to follow the changes in structure during the experiments. The GlpT WT structure is unfolded using urea, GuHCl or temperature. The reversibility of the unfolding reaction is assayed by recovery of the secondary structure upon reduction of the denaturant by dilution. The return of the tertiary structure and the functionality of the refolded GlpT is assessed with a substrate binding assay.

Finally, the effect of the ligand binding on the protein stability is established. In the end, the suitability of single tryptophan constructs as well as potential placements of rhodamine labels is discussed. The initial stability studies of single-Trp variants and the rhodamine label GlpT were made using circular dichroism and fluorescence. Firstly, the stability of single-tryptophans is evaluated. Followed by an analysis of the rhodamine labelled GlpT unfolding and its purification. Throughout the text the unlabelled GlpT construct is referred to as the cysteine mutant. When a GlpT rhodamine label is added it is referred to as a rhodamine labelled GlpT.

The outlined experiments and results are used in establishing the difference in stability of helices 1 and 7 in the following chapter 4 and possible future work in a summary chapter 5.

3.2 Results

3.2.1 Denaturant-induced unfolding at steady-state

The GlpT WT samples solubilised in 0.05% DDM were denatured with either 8 M urea or 6 M GuHCl. No further reduction of the secondary strucutre was observed after 5 minutes of incubation with either of these denaturants. The extent of the unfolding was measured by far-UV CD spectroscopy. The fluorescence measurements of GlpT's native tryptophans were also recorded.

The resulting changes in GlpT's CD or fluorescence spectra were plotted against the denaturant concentration to yield an unfolding curve, which in turn is used to calculate the free energy of unfolding in the absence of denaturant, provided that the reaction has reached equilibrium and is reversible.

3.2.2 Fluorescence measurements of unfolding in urea

A sample of GlpT WT was incubated in a range of 0 – 8 M urea concentrations for 5 minutes before fluorescence spectra measurements were taken. As the structure of GlpT becomes more denatured, with increasing urea concentrations, the environment around aromatic residues changes, resulting in a different fluorescence emission spectra for GlpT. Representative fluorescence emission spectra are shown in figure 3.4. The shift of the fluorescence emission spectra is coupled with a decrease in fluorescence intensity, which could be caused by the surrounding solvent quenching the fluorescence of the affected amino acid resides (160, 165, and 171). Between 0 M and 8 M urea concentration, a total of 48 % decrease in fluorescence intensity was observed with 280 nm excitation wavelength in regards to a 37 % reduction measured at 295 nm excitation wavelength. The shift of the wavelength of maximum fluorescence intensity was measured and plotted against urea concentration shown in figure 3.5. The observed red shift of the emission band is an indication of the environment around aromatic residues

becoming more polar. It does not directly relate to protein unfolding, but instead describes the increasing exposure of protein to the surrounding solvent.

The shift of the wavelength maximum against the urea concentration, for both excitation wavelengths, was fitted to the two state function with calculated midpoint of 4.9 ± 0.1 M and slope of 0.9 ± 0.1 for the 280 nm measurements and midpoint of 5.1 ± 0.1 M and slope of 0.9 ± 0.1 for the 295 nm excitation data, displayed in table 3.1. The 280 nm experiment spectra are due to fluorescence of tyrosine and tryptophan residues. Alternatively 295 nm or higher wavelengths are largely due to fluorescence of tryptophans. The discrepancy between different fluorescence experiments suggests that a different technique, which represents a more global change in the protein state would be more suitable to follow the GlpT unfolding. The choice of rhodamine labels and CD methodology is later discussed in this chapter and chapter 4 respectively.

The free energy change of unfolding in the absence of a denaturant (ΔG_U^{H2O}) was calculated from the emission band shift as a result of protein unfolding. The calculated unfolding slope and the midpoint obtained from the sigmoidal plot is equal to 4.5 ± 0.5 kcal.mol^{-1} and 4.8 ± 0.3 kcal.mol^{-1} for 280 and 295 nm excitation data, respectively. The ΔG_U^{H2O} was also calculated by a linear extrapolation method, where the ΔG values at each denaturant concentration were fitted to a linear equation. The y-intercept of the fitted straight line is descriptive of the ΔG_U^{H2O}. The linear fit for wavelength shift plotted against urea concentration can be seen in figure 3.6. The ΔG_U^{H2O} values derived from the linear fit are 3.4 ± 0.1 kcal.mol^{-1} for 280 nm excitation data set and 3.5 ± 0.1 kcal.mol^{-1} for 295 nm data set, with calculated slopes of -0.8 ± 0.3 and -0.7 ± 0.1 for 280 and 295 nm, respectively. The summary of results from fluorescence measured experiments of unfolding assays can be found in table 3.1. The proof of reaction reversibility and refolding curves are shown in the later section 3.2.4.

Measurement	Midpoint from sigmoidal fit	m value from sigmoidal fit	ΔG_U^{H2O} calculated from the sigmoidal fit (kcal.mol^{-1}.M)	ΔG_U^{H2O} calculated from the linear fit (kcal.mol^{-1})	m value from linear fit (kcal.mol^{-1}.M)	Total fluorescence lost (%)	Total emission band shift (nm)
280 nm excited peak shift	4.9 ± 0.1	0.9 ± 0.1	4.5 ± 0.5	3.4 ± 0.1	- 0.8 ± 0.3	48	2.1
295 nm excited peak shift	5.1 ± 0.1	0.9 ± 0.1	4.8 ± 0.3	3.5 ± 0.1	- 0.7 ± 0.1	37	3.1

Table 3.1 Summary of derived values from fluorescence urea unfolding assay

The ΔG_U^{H2O} with their midpoint and m value parameters were calculated from 280 nm and 295 nm urea unfolding experiments. The total fluorescence lost and total emission band shift were calculated from fluorescence spectra shown in figure 3.4. The errors are taken from standard error of the sigmoidal or linear fits of unfolding assays, where appropriate.

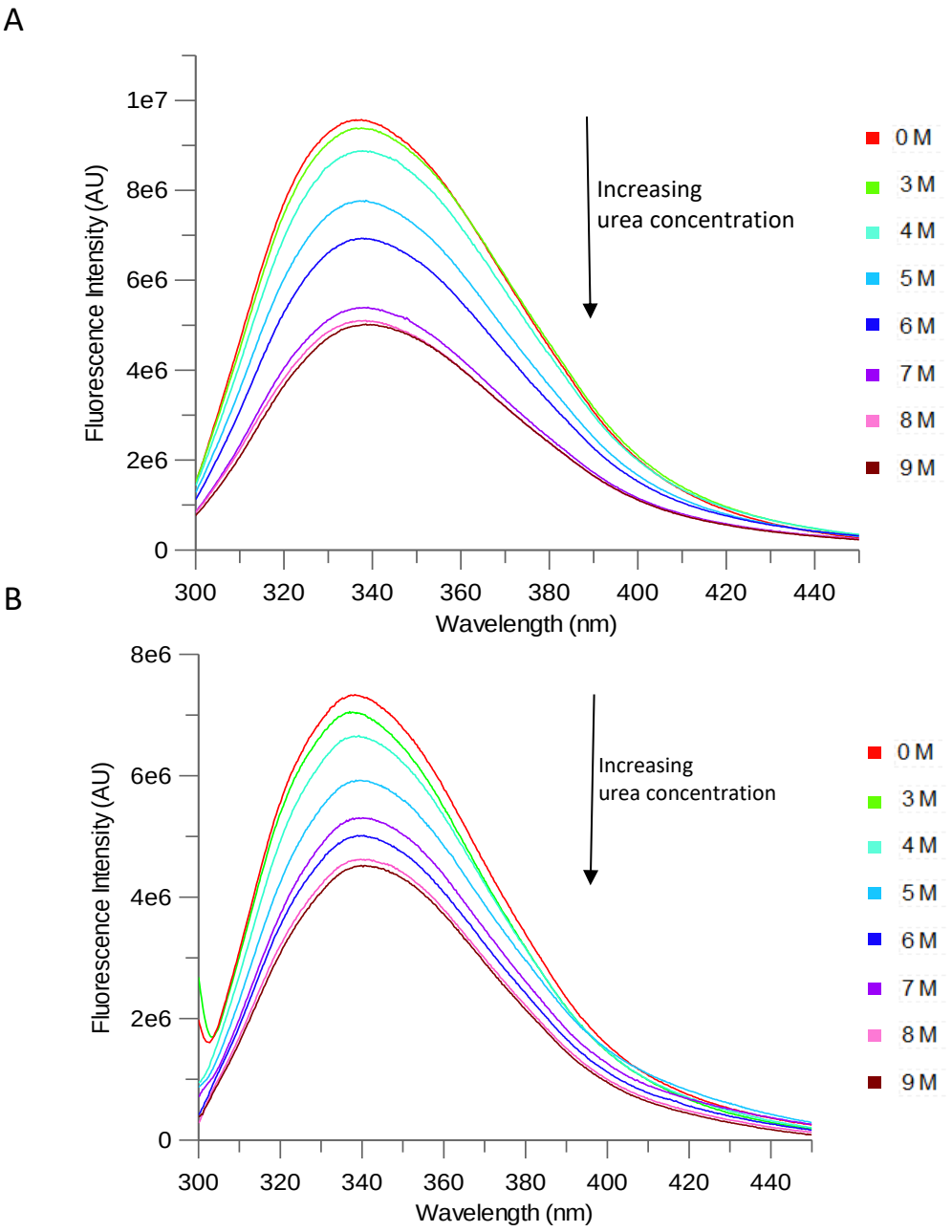


Figure 3.4 Representative fluorescence spectra of GlpT WT incubated with urea

Representative spectra of 0.025 mg.ml^{-1} GlpT WT, in a 4 mm pathlength quartz cuvette, measured at 280 nm and 295 nm excitation wavelength (in panels A and B, respectively), at increasing concentrations of urea displayed in different colours on the right. Decrease in fluorescence can be seen as the concentration of urea increases. The largest decrease is seen around the 5 M mid-point. A decrease of the fluorescence intensity due to urea unfolding is seen at both excitation wavelengths and is equal to 48 % decrease of the original fluorescence intensity for 280 nm excitation and 37 % for 295 nm excitation wavelength.

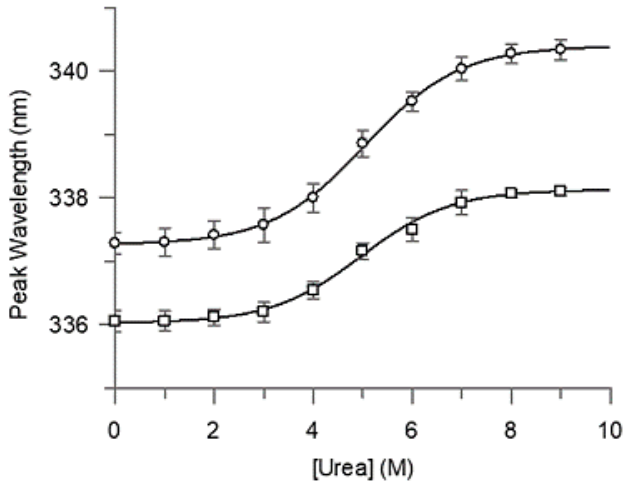


Figure 3.5 Wavelength shift recorded upon urea unfolding with GlpT WT

Measurements at 280 nm excitation wavelength, shown as squares produce a smaller wavelength shift of 2.1 nm compared to measurements at 295 nm, represented by circles which produce a 3.1 nm shift. All spectra were measured with 0.025 $mg.ml^{-1}$ of GlpT WT in a 4 mm pathlength quartz cuvette, incubated in urea for 5 minutes with data collected at 0.25 nm increments from between 300 and 450 nm with a slit width of 5 nm.

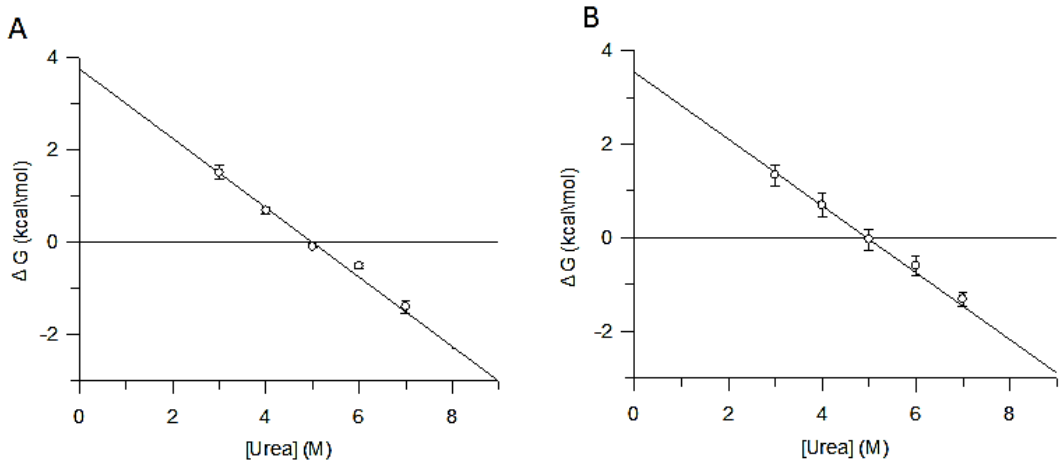


Figure 3.6 Linear fits of GlpT WT unfolding measured by fluorescence

GlpT WT was unfolded in urea and the shift of wavelength at maximal fluorescence was recorded to monitor the state of the protein. The linear transformation of 280 and 295 nm excitation data sets can be seen in panels A and B, respectively. The fit y intercept, equivalent to ΔG_U^{H2O} was calculated at 3.4 ± 0.1 $kcal.mol^{-1}$ for 280 nm excitation data set and 3.5 ± 0.1 $kcal.mol^{-1}$ for 295 nm data set.

3.2.3 CD measurements of unfolding in urea

GlpT WT unfolding was monitored by a reduction in structure measured by CD spectroscopy. Samples were incubated in a 0 – 8 M range of urea concentrations, at 1 M intervals, for 5 minutes at room temperature. The full recorded spectra, between 180 to 260 nm, of folded GlpT WT in a DDM micelle with no urea present is shown in figure 3.7 at 0 M urea concentration. The folded spectra were fitted to the DICHROWEB (147) database, which predicted the GlpT sample to be composed of 75% alpha helices. This is in close agreement with previous studies (93, 122) and the published GlpT crystal structure of 71 % (98). Approximately 33% of the GlpT's structure is lost in 8 M urea. An additional information on the state of the protein can be measured from recording a positive 190 nm band, which has an intensity strongly dependent on protein helicity and is therefore required for secondary structure content estimation. The high absorption of urea below 210 nm wavelength makes the measurement inaccessible, hence all subsequent scans at increasing urea concentration were taken between 205 and 260 nm. The unfolded GlpT WT CD spectra is shown in figure 3.7. The CD scans were converted into a normalised value of mean residual ellipticity (MRE) against which urea concentrations were plotted to yield the unfolding curve of GlpT WT in a DDM micelle, seen in figure 3.8. The negative band at 222 nm wavelength was used as an indicator of structure lost, as well as to estimate the percentage of protein unfolded.

The concentrations of 1 and 2 M urea do not denature GlpT, as determined by this assay. The majority of denaturation occurs above 3 M urea, with the midpoint of the denaturation curve occurring around 4.5 M urea. If the CD unfolding data is fitted to a sigmoidal plot the resulting parameters are 4.7 ± 0.1 M for midpoint value of 0.9 ± 0.1 for slope, shown in figure 3.8. The resulting free energy change calculated from the sigmoidal plot is equal to 4.3 kcal.mol^{-1}, compared to free energy change from linear plot, shown in panel B, indicated by the y intercept at 3.8 kcal.mol^{-1}. These fits and resulting free energy values are made under an assumption that the unfolding reaction is reversible and follows a two state model for protein folding. The assumption is later tested in section 3.2.4.

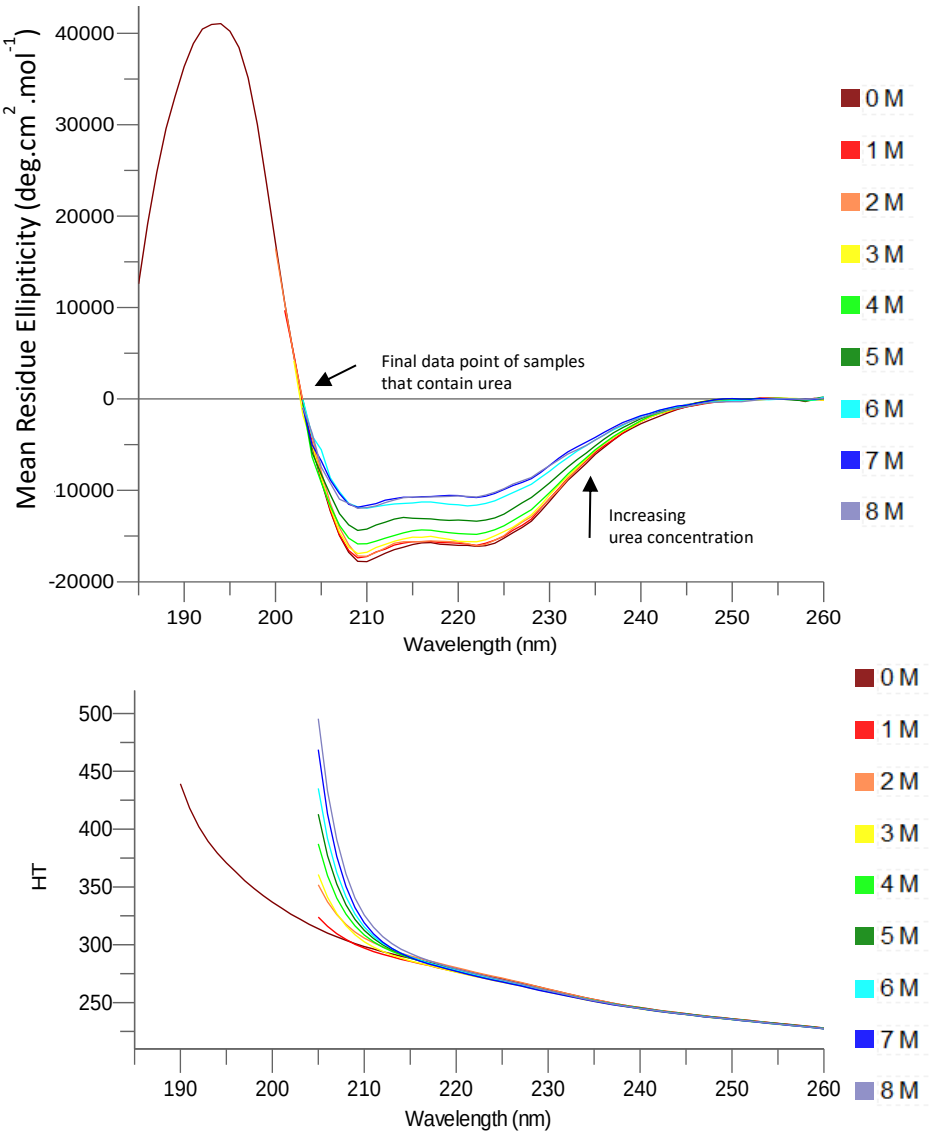


Figure 3.7 CD spectra of folded GlpT WT and incubated in urea

In top figure, GlpT WT was unfolded in increasing concentrations of urea for 5 minutes at 25°C after which CD spectra was taken to follow a decrease in the negative band centred at 222 nm wavelength. The decay of MRE at 222 nm is plotted in figure 3.6 and fitted to sigmoidal or linear equation to yield the ΔG_U^{H2O} value. Spectra were recorded in a 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 260 nm to 180 nm in buffer and from 260 nm to 205 nm in denaturant containing samples. HT values for each urea concentration sample are shown in the bottom figure. In each case HT value did not exceed 500, therefore allowing use of scans for qualitative analysis. The raw spectra from three repeated measurements was first averaged with the buffer scan subtracted, the baseline between 260 and 255 nm was zeroed in CDTool then plotted and converted to MRE in GraFit software.

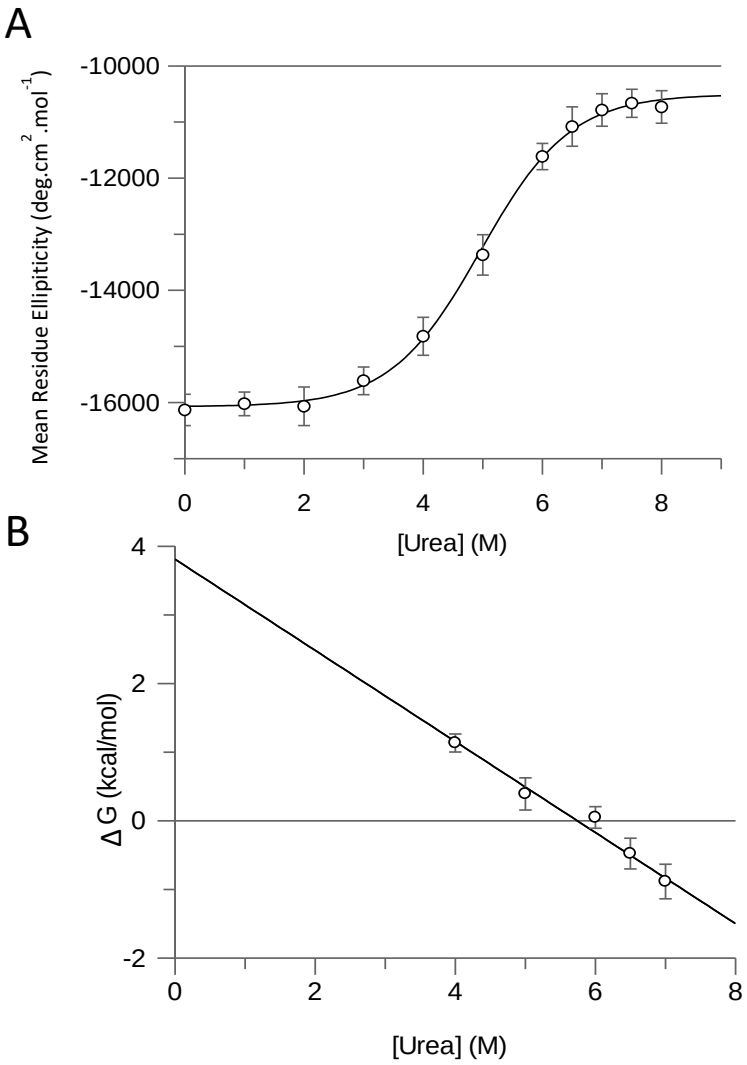


Figure 3.8 Unfolding of GlpT WT in urea measured by CD fitted to sigmoidal plot

GlpT WT was unfolded in urea for 5 minutes and the degree of unfolding monitored by a decay of the 222 nm band intensity was plotted against urea concentration. The resulting data was fit to a sigmoidal plot with a midpoint at 4.6 ± 0.1 M and slope equal to 0.9 ± 0.01, with errors from the standard fit of the plot, shown in panel A. Alternative method of data fitting is shown in panel B with the y intercept of the linear fit giving 3.8 ± 0.4 kcal.mol^{-1} and gradient equal to - 0.8 ± 0.1.

3.2.4 Reversibility of the urea unfolded GlpT WT

The ΔG_U^{H2O} calculation for a two state unfolding reaction, shown in sections 3.2.2 and 3.2.3, is based on the assumption that the unfolding reaction exists in a reversible equilibrium and consists of only two different states, unfolded and folded.

The return of the secondary structure from the urea induced unfolded state to the original folded state was assessed by far-UV CD. The refolding reaction was initiated by rapid dilution of the urea. The spectra for the folded, unfolded and refolded states can be seen in figure 3.9. The spectrum of the refolded protein overlays with spectrum of the native GlpT before denaturation. It is observed that both bands at 222 and 208 nm overlay, signifying the return of the original intensity of CD signal when the concentration of denaturant is reduced ten-fold. For comparison, the unfolded spectra of GlpT is shown in red. Measuring the CD spectra of a refolded protein is difficult as the large dilution of the protein during the refolding results in a decreased CD signal. The use of higher pathlength cells and SRCD high intensity beam helps to reduce the noise found in the measurements. Moreover, the accuracy of dilution affects the total volume calculation and thus final protein concentration, which in turn produces additional errors during the CD signal to MRE conversion.

The refolding reaction was repeated with GlpT being refolded using a range of refolding buffers containing different concentrations of urea. The resulting refolding curve, where the return of a CD signal, present at 0 M urea concentration, can be followed at each urea concentration is shown figure 3.10. The refolding curve shows an identical shape to the unfolded curve presented in the same figure for comparison. Although CD measures return of the structure, it does not provide detail into assembly of tertiary contacts. An additional ligand binding assay that supplements the CD data is described in section 3.2.9.

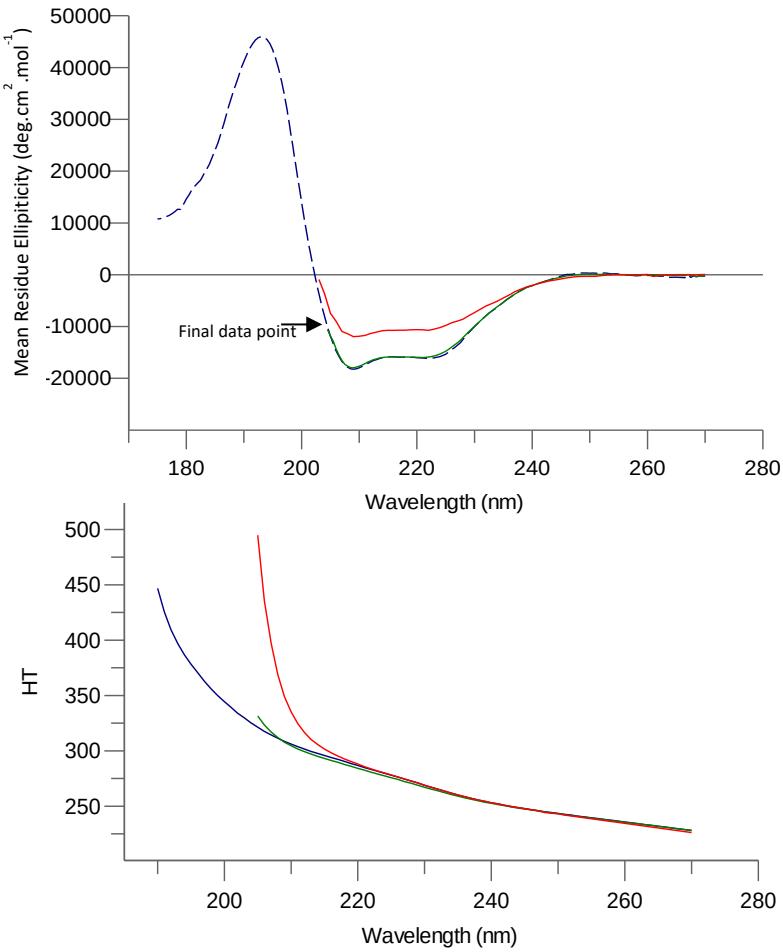


Figure 3.9 CD spectrum of refolded GlpT compared with the folded and unfolded sample

GlpT WT was refolded by a ten-fold dilution of GlpT sample incubated with 8 M urea. Top figure shows the overlay of refolded and folded samples, shown here in green and blue respectively. Similarity between folded and refolded CD scans represents the return to similar conditions like in the original structure before the denaturation by urea. The protein was unfolded for 5 minutes in 8 M urea at room temperature giving the spectra seen in red. The sample was then refolded by a ten-fold dilution into buffer and incubated for 10 minutes at room temperature before taking the full spectrum. HT values for each urea concentration sample are shown in the bottom figure. In each case HT value did not exceed 500, therefore allowing use of scans for qualitative analysis. The SRCD scans of folded and refolded samples were recorded at ANKA with 12 µm and 0.5 mm pathlength cells, respectively, with 0.5 wavelength mm intervals between 175 – 270 nm for folded protein and 200-270 nm for refolded sample. The unfolded spectra was measured using in-house Aviv Circular Dichroism Spectrophotometer, Model 410 in 0.2 mm pathlength with 1 mm interval, the spectra shown are an average of 3 repeated experiments.

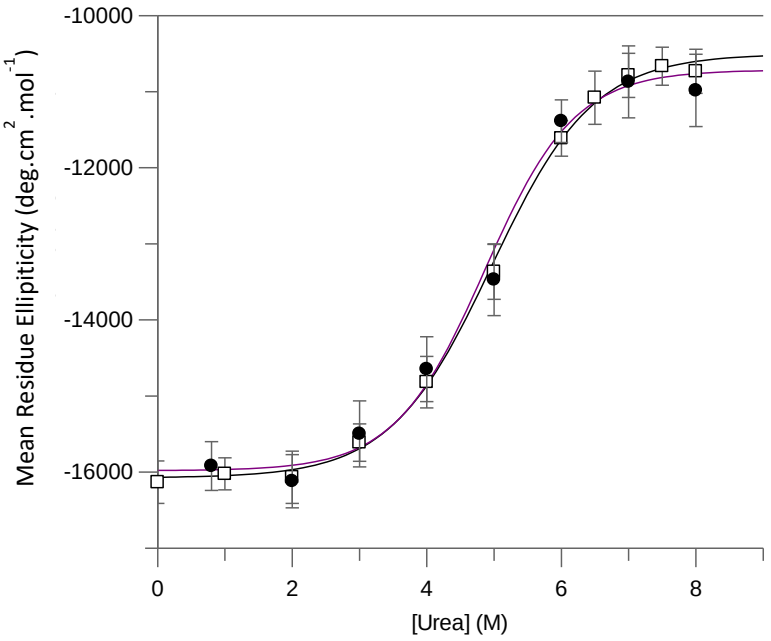


Figure 3.10 Comparison of GlpT WT unfolding and refolding curves measured by CD

The extent of GlpT structure change was monitored by CD signal at 222 nm band for both unfolded sample, shown as open squares, and refolded sample, shown as black circles, the errors are from standard error of the fit. The CD signal measured at 222 nm was converted to MRE and overlaid to demonstrate similarities between unfolding and refolding processes. The data was recorded with 0.2 mm and 0.5 pathlength cell for unfolded and refolded samples, respectively, at 1 nm intervals and 1 second acquisition time from 260 nm to 200 nm. The raw spectra were first processed in CDTool then converted to MRE and plotted in GraFit software.

3.2.5 Fluorescence measured unfolding induced by GuHCl

Alpha helical membrane proteins are naturally resistant to denaturation and do not unfold completely, due to high energy cost of exposing their hydrophobic core (18). To further test GlpT stability an additional unfolding assay with GuHCl was performed. The data presented is the result of a single experiment and therefore needs to be repeated for more accurate analysis. The use of a chaotrope, GuHCl, increases the degree of protein unfolding, resulting in a greater loss of secondary strucutre and an increased loss of fluorescence upon incubation with the denaturant, due to greater degree of aromatic residues exposed to surrounding solvent and therefore fluorescence quenching. The change of protein fluorescence was monitored by excitation at 295 nm wavelength. In

figure 3.11, the fluorescence emission spectra of GlpT WT denatured with various concentrations of GuHCl are presented. The GuHCl incubation shows a more drastic effect on wavelength shift at fluorescence maximum as well as maximal fluorescence intensity compared to the denaturation experiment with urea.

Approximately 37 % of fluorescence is lost in the 295 nm excitation experiment. In the same way as in urea denaturation assay, a shift of the wavelength maximum was monitored and estimated from the normal distribution fit. The resulting wavelength shift data can be seen in figure 3.11, with an overall 5.5 nm shift for the 295 nm excitation measurement. When compared to urea denaturation, GuHCl shows an increased red shift of the fluorescence spectra when measured in 295 nm excitation wavelength. Fluorescence measurements of GuHCl induced unfolding show different unfolding behaviour with different end points due to different denaturants having different mechanisms of denaturation.

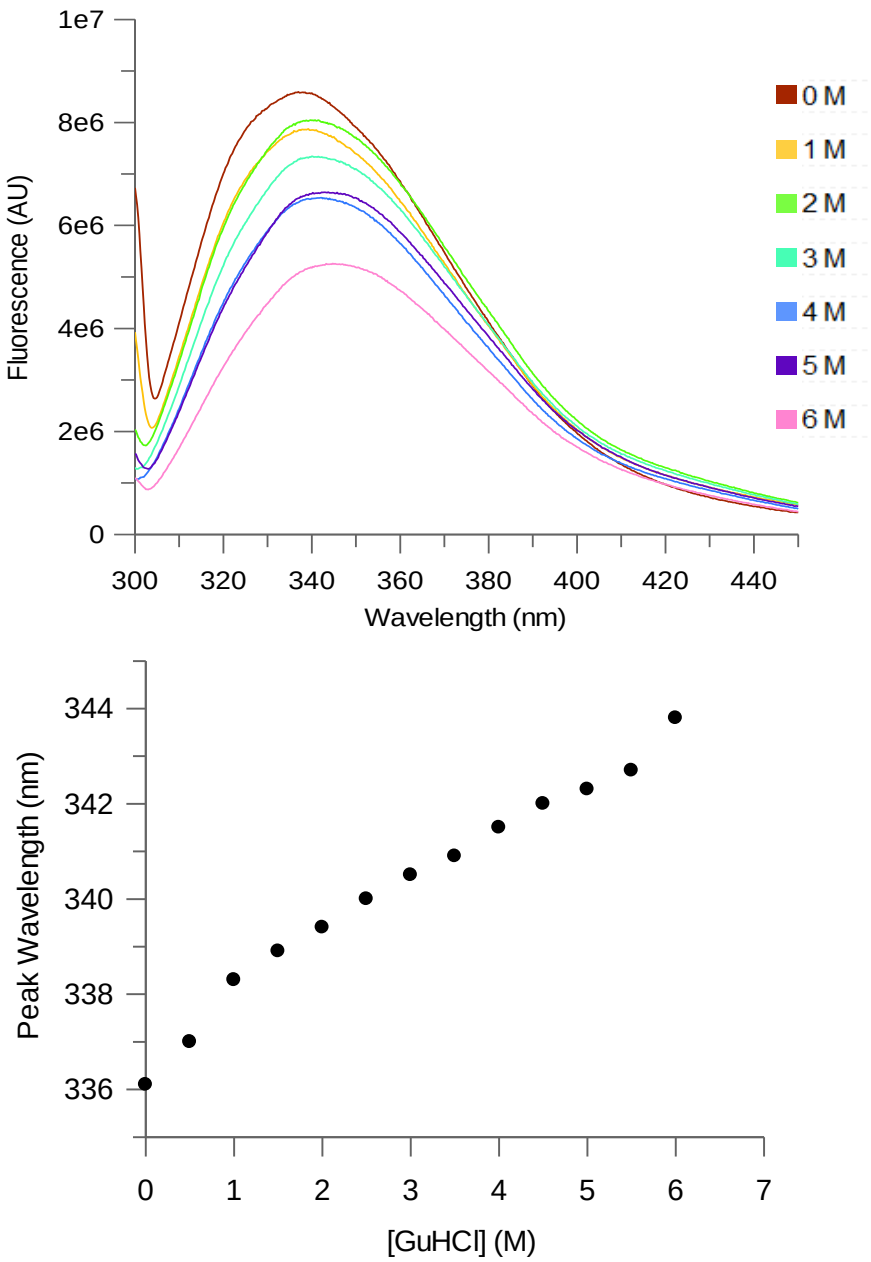


Figure 3.11 GlpT WT unfolding in GuHCl measured by fluorescence

GlpT WT was unfolded in increasing concentrations of GuHCl for 5 minutes before spectra were recorded at 295 nm excitation wavelength. Different concentrations of GuHCl are represented by their respective colour on the right, top figure. The decrease and shift of the fluorescence peak can be noted, with an overall decrease of fluorescence intensity of 37 % and overall wavelength shift of 5.5 nm for 295 nm measurement. The peak wavelength data was analysed and plotted in the bottom figure, showing the wavelength shift at increasing GuHCl concentrations. All spectra were measured with 0.025 $mg.ml^{-1}$ of GlpT WT in a 4 mm pathlength quartz cuvette, incubated in urea for 5 minutes, with data collected at 0.25 nm increments from between 300 and 450 nm with a slit width of 5 nm.

3.2.6 CD measured unfolding induced by GuHCl

The change in GlpT's structure induced by a 5 minute denaturation with GuHCl denaturant was measured *via* CD spectroscopy. The resulting spectra of different steps of denaturation can be seen in figure 3.12. The overall loss of CD signal due to GuHCl is reduced by around 57 %, compared to a loss of 33 % in 8 M urea.

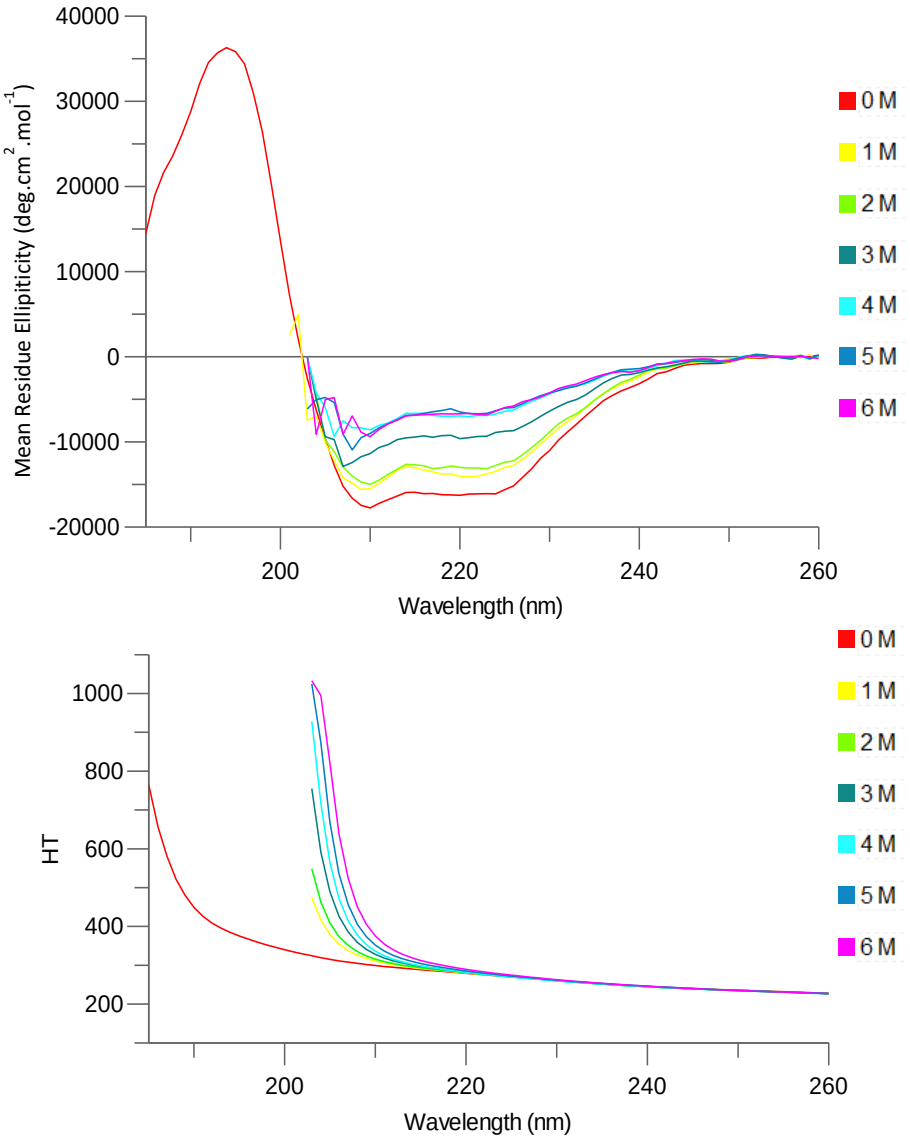


Figure 3.12 CD spectra of GlpT WT incubated in GuHCl

Top figure, GlpT incubated with increasing concentrations of GuHCl for 5 minutes, before the CD spectra were taken. Different concentrations of GuHCl are represented by their respective colour on the right. The intensity of the band at 222 nm decreases by 57 %, showing more dramatic loss of CD signal when compared to urea denaturation. HT values for each GuHCl concentration sample are shown in the bottom figure. In each case HT values recorded below 210 nm did not

exceed 500, therefore allowing use of scans for qualitative analysis. Spectra were recorded with 0.025 mg.ml^{-1} of GlpT WT in a 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 260 nm to 180 nm in buffer and from 260 nm to 205 nm in denaturant containing samples. The raw spectra were first processed in CDTool then plotted and converted to MRE in GraFit software.

The loss of CD signal was monitored by decay of the 222 nm band and plotted against increasing concentrations of GuHCl shown in figure 3.13. The data presented is the result of single scans and therefore needs to be repeated for a more accurate analysis.

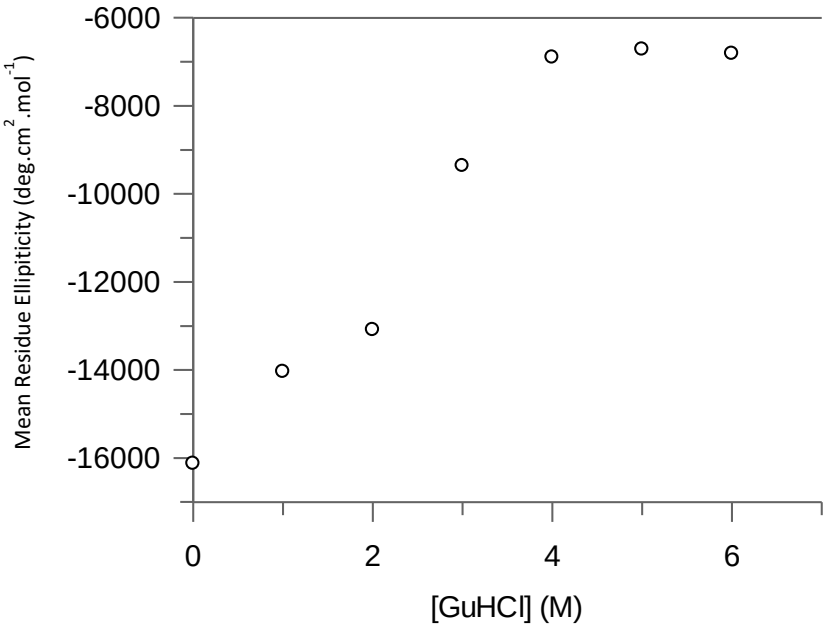


Figure 3.13 Unfolding of GlpT WT in GuHCl measured by CD

The intensity of the 222 nm band was plotted against concentrations of GuHCl to follow the degree of unfolding of the protein. There is an immediate loss of structure due to GuHCl with steep pre-transition baseline and noticeable saturation at 4 M concentration. The data was first processed in CDTool then 222 nm band was converted to MRE and plotted against GuHCl concentrations in GraFit software.

3.2.7 Time-resolved measurements of unfolding

Urea induced unfolding of GlpT was further assessed by recording time-resolved measurements of fluorescence intensity during the process of protein unfolding. The goal was to establish a time-resolved experiment to show GlpT's two state behaviour during urea unfolding, as well as to gain further insight into protein stability. The experiments were performed on protein purified using a gel filtration column as a last step of purification in order to avoid any residual dimer unfolding that may cause incorrect interpretation of rate constants, thus affecting results. The loss of fluorescence intensity, associated with the unfolding may be misinterpreted when recording the kinetics of the dimer unfolding due to an extra step in the unfolding reaction. An extra step is possibly linked to the dissociation of the two monomers at low concentrations of the urea, followed by unfolding of more denaturation resistant individual monomers. The steady state spectra were not influenced by the final method of protein preparation since the measurements were taken after samples had reached equilibrium. GlpT was unfolded in urea and each unfolding experiment was recorded for the first 200 seconds with the first 10 seconds recorded at increased frequency in order to obtain more data points in the initial stages of the experiment. Up to eight traces were recorded for protein unfolding at the stated urea concentration. The traces were then averaged to reduce the noise, followed by fitting to a double exponential equation. Occasionally, a single exponential fit could have been made to unfolding performed at higher urea concentrations. However, it was not possible to fit any measurements to a single exponential fit with 6.5 M urea or below. To make all analysis comparable, all fits were made to a double exponential function. Representative unfolding measurements recorded during GlpT unfolding are shown in figure 3.14 with the residual for the double exponential fit to the 8 M trace presented below.

The two rate constants calculated from each fit, corresponding to denaturation with different urea concentrations, were converted to a natural logarithm and plotted against their corresponding urea concentration, shown in figure 3.15. The rate of unfolding in the

absence of the denaturant can then be calculated from the inverse of the natural logarithm defined by an intercept at the y-axis. The error of each point on the linear plot is the result of the difficulty to accurately fit an exponential rate of decay to the first few seconds of the reaction. The reason for the GlpT's unfolding behaviour at the beginning of the fluorescence recording could be due to a rich aromatic environment that is wide-spread throughout the protein and could have led to a complex fluorescence behaviour and artefacts during the initial seconds of the unfolding.

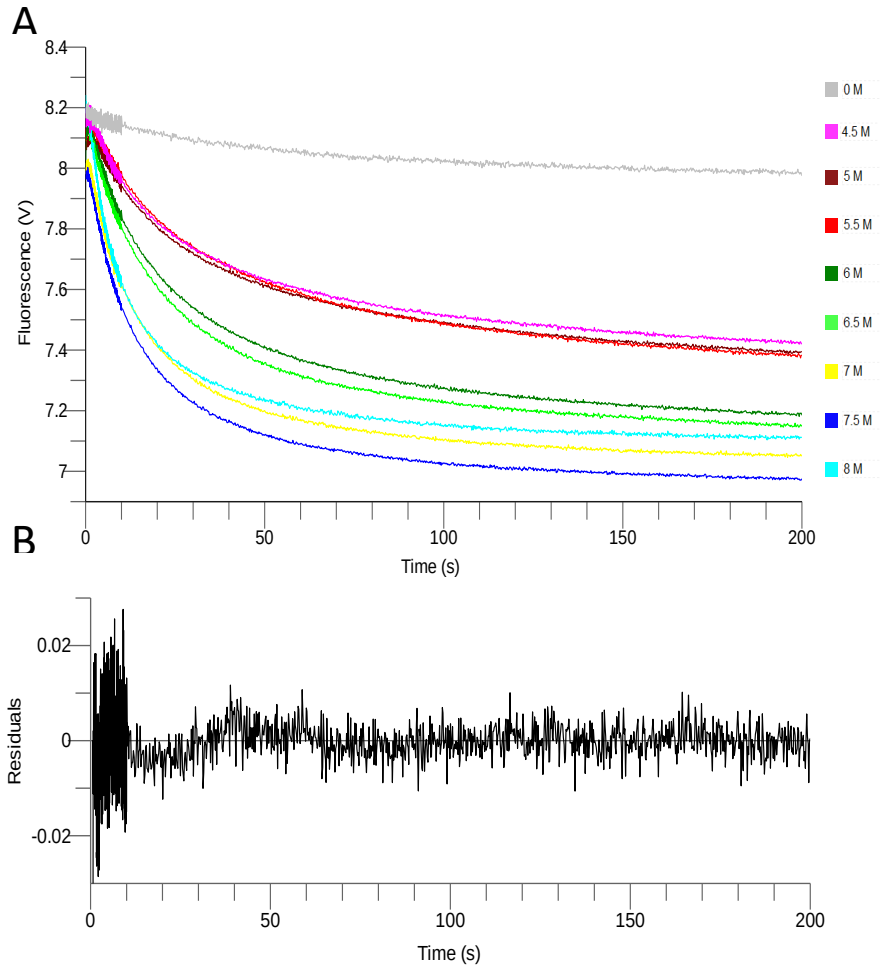


Figure 3.14 Time resolved unfolding of GlpT WT in urea measured by fluorescence

Traces of unfolding of GlpT are shown at individual urea concentrations, shown in panel A. Different concentrations of urea are represented by their respective colour on the right. The top trace at 0 M urea concentration was recorded with protein in buffer. It shows fluorescence decay due to photo bleaching. Each trace is an average of up to 8 repeats, before the resulting trace was fit to a double exponential fit. A single exponential equation could not be fit accurately, mainly due to the unclear initial rate measurements. An example of the residual from the fit is seen below in panel B.

The resulting first and second rate constants from the linear plot were calculated to be $6.7 \times 10^{-3} \pm 4.0 \times 10^{-3}$ s^{-1} and $3.9 \times 10^{-3} \pm 5.7 \times 10^{-4}$ s^{-1} respectively, with error taken from the standard error of each linear fit. The gradient of the linear fit to kinetic data is characteristic of the dependency of the unfolding on the denaturant concentration, with the first and second rate gradient equal to 0.42 ± 0.05 kcal.mol^{-1}.M and 0.22 ± 0.01 kcal.mol^{-1}.M, respectively. The second reaction rate could be associated with the protein undergoing photobleaching during the experiment (172). The GlpT photobleaching during this assay can be seen from the grey coloured trace, measured in the absence of urea, in figure 3.14. The lower dependency on the denaturant, described by a smaller gradient of second rate compared to first rate confirms this possibility. The unreliability of kinetic measurements makes the assessment of a two state behaviour difficult. Future experiments with time resolved measurements and single fluorescence probes discussed later should clarify and conclusively show the existence or lack of any intermediates in the GlpT's unfolding reaction mechanism.

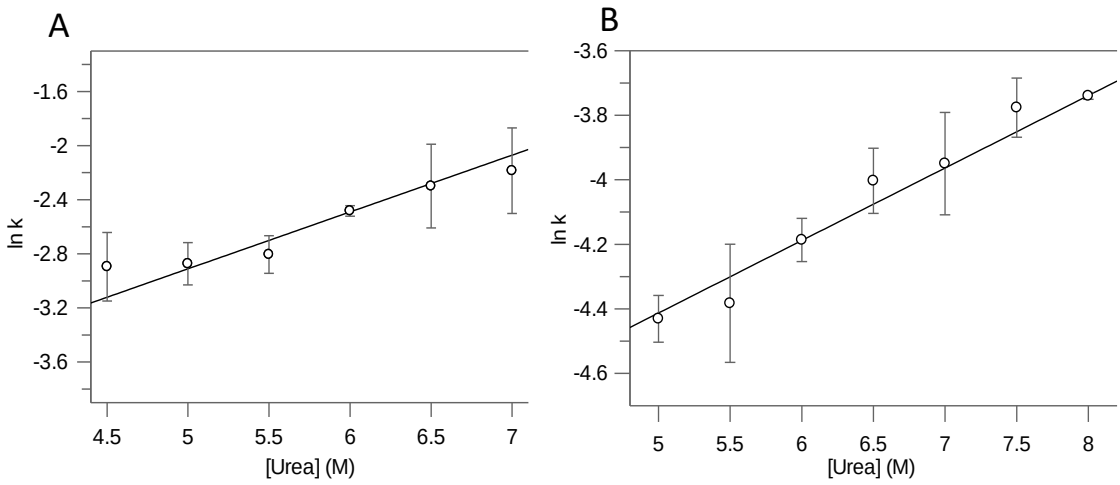


Figure 3.15 Unfolding kinetics of GlpT WT

The rate of unfolding for each urea concentration was calculated from the double exponential fit with the resulting first and second rate shown in panels A and B, respectively. The natural logarithm of the rate of unfolding in the absence of denaturant, represented by the intercept at y-axis was calculated to be -5.0 ± 0.3 and -5.5 ± 0.1 for first and second rate respectively, the errors shown are from standard error of the fit and a result of 4 repeats.

3.2.8 Thermal denaturation of GlpT WT

The GlpT unfolding was investigated with increasing temperature. At each temperature step a recording of either fluorescence or CD signal was taken in order to monitor the state of the protein, shown in Figure 3.16. The CD signal at 222 nm band, the maximum fluorescence intensity and the peak fluorescence wavelength were chosen as parameters to construct denaturation curves that were then fitted to sigmoidal equations, with the midpoint of the transition curve taken as the denaturation midpoint Tm. The temperature denaturation curves for fluorescence and CD measurements are found in figures 3.17 and 3.18, respectively. The concentration of the protein used in each temperature denaturation experiment was kept constant at 0.1 $mg.ml^{-1}$ in order to keep calculated Tm values comparable. The summary of individual Tm values from each parameter recorded can be seen in Table 3.2

Parameter recorded	Tm (°C)
Fluorescence intensity (280 nm excitation)	51.0 ± 0.3
Fluorescence intensity (295 nm excitation)	54.6 ± 0.5
Wavelength at fluorescence emission maximum (280 nm excitation)	52.9 ± 1.3
Wavelength at fluorescence emission maximum (295 nm excitation)	53.2 ± 1.1
CD signal decay at 222 nm	49.0 ± 0.2

Table 3.2 Thermal denaturation of GlpT WT

The Tm values of GlpT WT were calculated by fitting sigmoidal equation to each denaturation curve, the errors are given from the standard error of the fit. Regardless of the method of analysing the temperature denaturation in fluorescence based experiments, the resulting Tm is in the same range of 53 °C. Results from a CD measurement lead to a Tm value with of 49.0 °C.

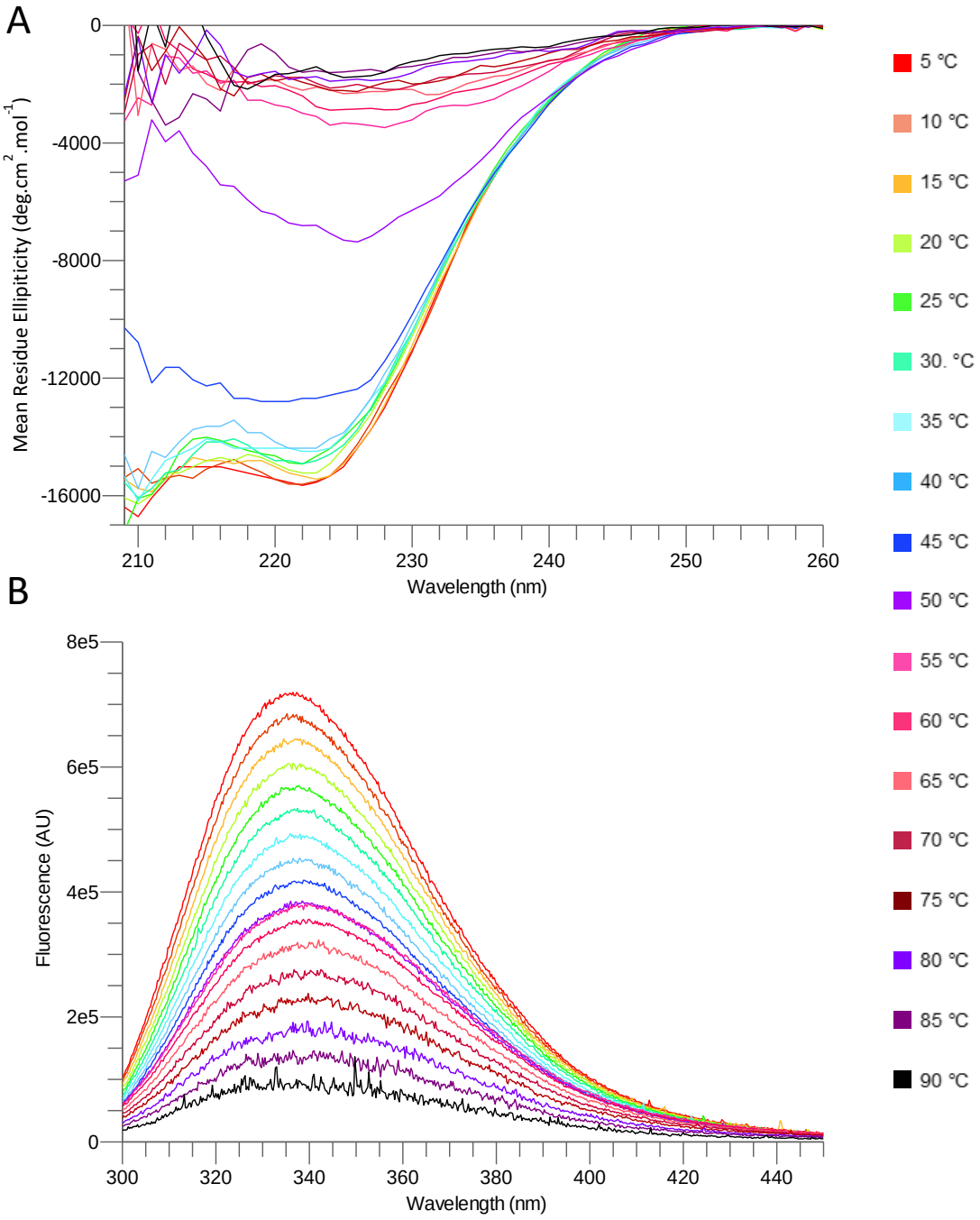


Figure 3.16 Thermal denaturation scans of GlpT WT

Each scan was taken at an increasing different temperature on the same protein sample. The corresponding temperature of each scan is displayed on the right for both A, CD measured denaturation, and B, fluorescence measured denaturation. Fluorescence spectra were recorded with 0.1 mg.ml^{-1} GlpT WT, in a 1 mm pathlength quartz cuvette measured at 280 nm. The CD spectra were recorded with 0.1 mg.ml^{-1} of GlpT WT in a 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 260 nm to 180 nm in buffer.

Overall, there is no difference between Tm values determined from fluorescence intensity change or wavelength maximum, as most Tm values lie close within 53 °C range. The protein unfolding measured by CD has a smaller pre-transition slope and bigger transition phase than the fluorescence recorded experiment. The resulting Tm from CD temperature denaturation was determined to be 49.0 ± 0.2 °C. Different transition phases and resulting Tm values between fluorescence and CD temperature experiments could be due to a CD signal change being a global representation of shifting structure content and therefore a more accurate description of protein unfolding.

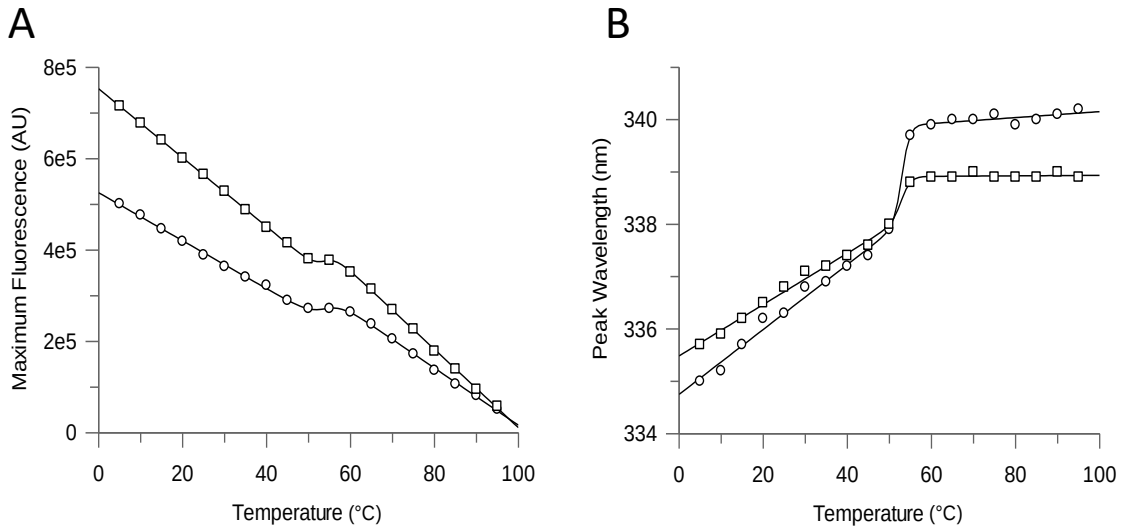


Figure 3.17 Thermal denaturation of WT measured by fluorescence

The fluorescence scans, shown in figure 3.16, were taken at 5 °C intervals, each scan was fitted with a normal distribution fit to estimate maximal fluorescence intensity (panel A) and fluorescence emission maximum (panel B), presented in the figure above. A sigmoidal equation with fixed pre- and post-transition slopes was fit to both experiments at 280 nm excitation, shown as squares, and 295 nm excitation, shown as circles, to find the midpoint of transition upon thermal denaturation, Tm. The summary of Tm values from different experiments is given in table 3.2.

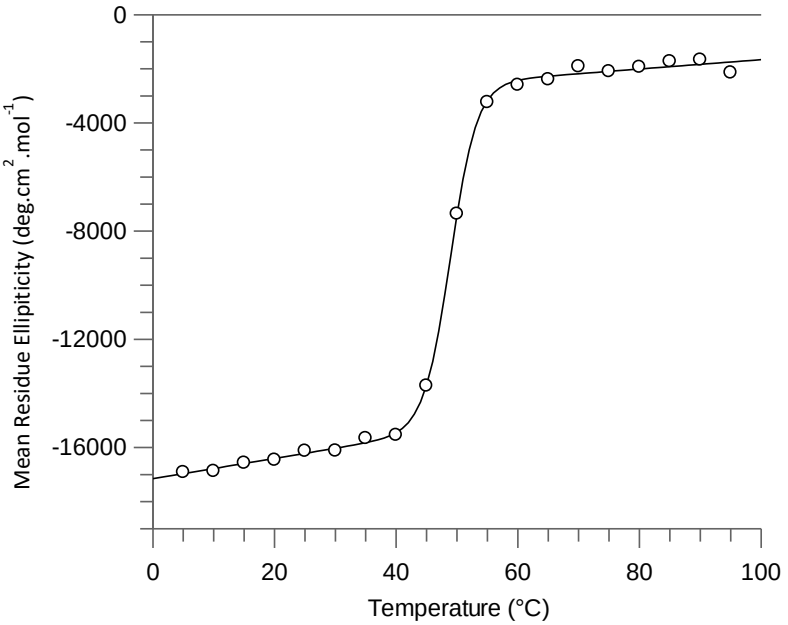


Figure 3.18 Thermal denaturation of WT measured by CD

The CD scans were taken at 5 °C intervals to monitor the extent of CD signal loss with increasing temperature. The decay of CD signal at 222 nm was then fit to sigmoidal curve to find the midpoint of transition upon thermal denaturation, Tm of 49.0 ± 0.2 °C, with the error taken from the standard error of the fit.

3.2.9 Ligand binding

A ligand binding assay was employed as an additional assessment of the efficiency of GlpT to undergo refolding from the urea denatured state. The CD refolding assay discussed in section 3.2.4 shows the return of pre-denaturation CD spectra, however it does not provide any extra information on the return of tertiary structure content. By comparing the binding pocket's affinity for the ligand before and after refolding it is possible to assess whether the refolding is fully reversible.

During the protein refolding, both the protein and denaturant are diluted tenfold from the saturating urea concentration of 8 M to a concentration of 0.8 M. The unfolding assay, discussed in section 3.2.3 shows the majority of denaturation occurs above 3 M urea.

Therefore, the structured content present at 0.8 M urea should closely resemble that of the non-denatured GlpT. Furthermore, a tenfold dilution of the sample is desirable in this experimental design, in order to avoid saturating binding sites at low ligand concentrations, thus enabling accurate measurement of the ligand binding constant. If the protein concentration exceeds the binding constant it can lead to a false recording of ligand saturation and therefore incorrect calculation of the dissociation constant.

The binding affinity in transporters is usually found in the µM to mM range. Several methods and compounds were used to find an appropriate ligand with strong binding affinity in the µM range, which would enable easier estimation of the return of substrate affinity upon protein refolding. Furthermore, the ligand binding needs to be accurately and robustly measured in the detergent micelle-protein system. Through literature research, 4 ligands were identified for their strong affinity to GlpT: G3P, P_i, P5P and phosphomycin. Fluorescence quenching *via* ligand binding, presented in the previous research (98) was found to be difficult to repeat due to the small degree of fluorescence quenching and photobleaching during the titration experiments with the same sample.

The ligand binding assay was also measured by recording changes in the CD signal upon ligand binding. Out of 4 ligands, P5P and phosphomycin showed the largest change in the CD spectra with a 7 % and 3 % signal increase at the 222 nm band, respectively. The SRCD spectra recorded with GlpT WT incubated with phosphomycin and P5P are shown in Figure 3.19. Deconvolution of CD spectra of the GlpT WT with phosphomycin and P5P using the SMP180 dataset gives helical structures of 70 % and 78 % alpha helix for phosphomycin and P5P bound GlpT WT, respectively. Both G3P and inorganic phosphate had no effect on CD spectra.

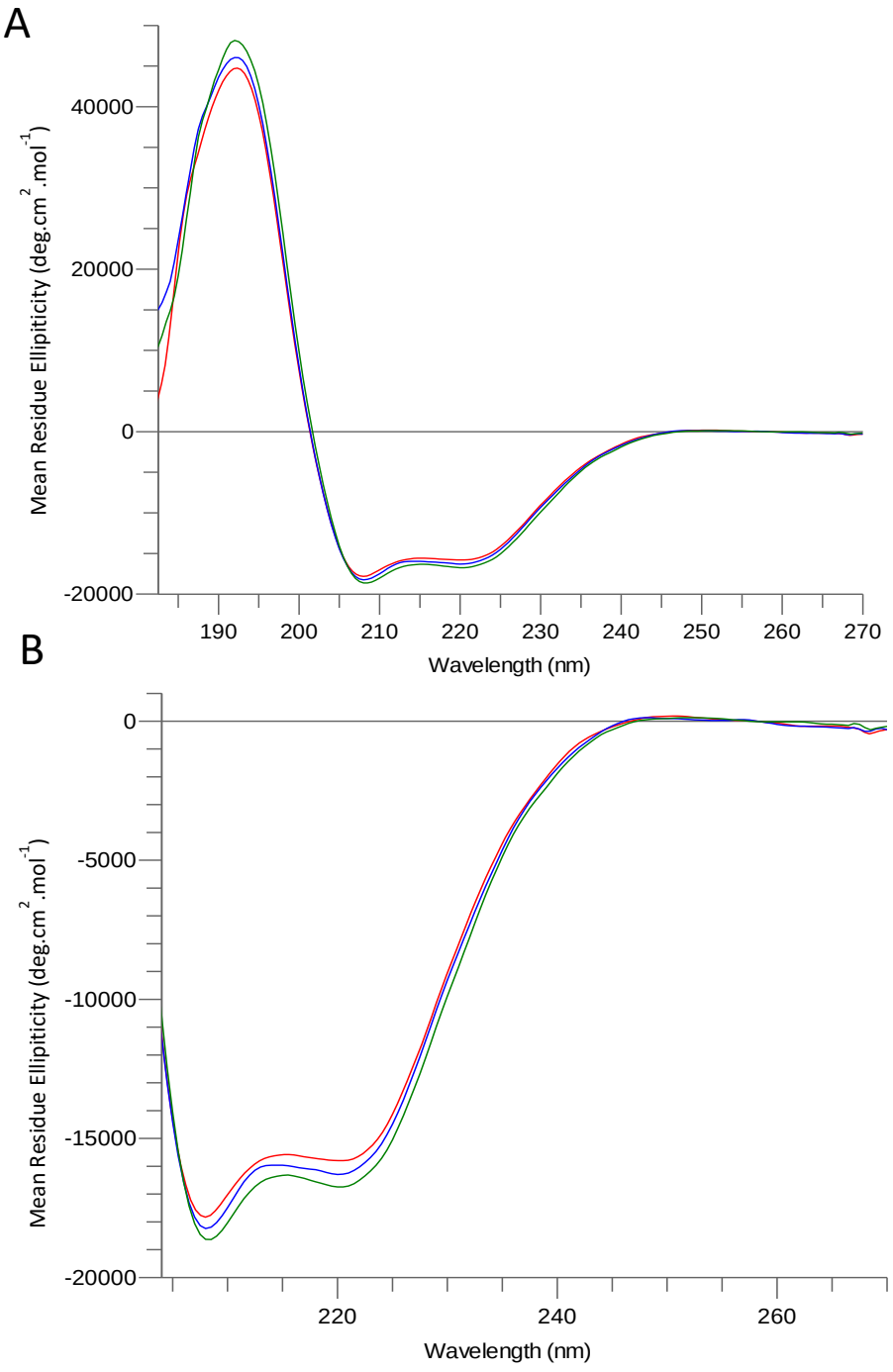


Figure 3.19 CD scan of WT incubated with ligands

GlpT WT was incubated for 10 minutes with phosphomycin, shown in blue or P5P represented by green line and GlpT WT with no ligand in red. Spectra were recorded with 0.1 mm pathlength cells at ANKA SRCD source at 0.5 nm intervals and 1500 ms averaging time from 180-270 nm. Panel A shows whole spectra recorded, negative alpha helical bands of the same measurements can be seen magnified in panel B. The raw spectra were processed with CDTool and plotted using GraFit software.

The P5P was chosen as the best ligand for the refolded ligand binding assay due to its large effect on the CD signal change. A 2 mm pathlength cell was used in recording CD spectra due to the low protein concentration necessary for ligand binding. Precautions were made to make sure that bigger pathlength does not lead to any light scattering during the experiments and therefore contribute to a smaller change of CD signal. CD signal change at 222 nm was used to monitor the change, and was plotted against P5P to give saturation binding curve shown in figure 3.20. Compared to the binding assay recorded with fluorescence, the CD orientated method provided a more accurate and repeatable method of calculating Kd.

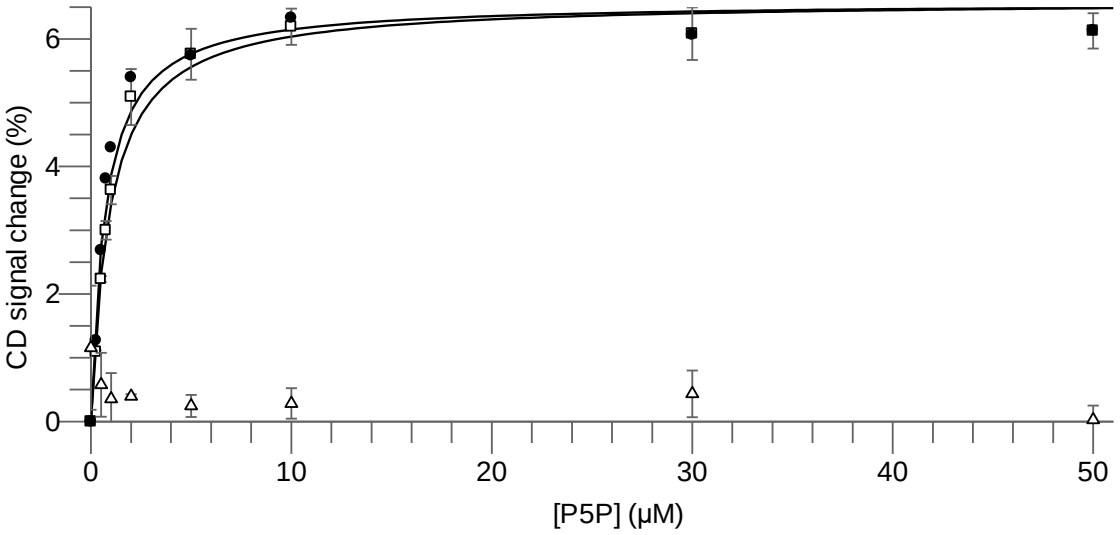


Figure 3.20 Saturation binding experiment with GlpT WT incubated with P5P

Changes in CD signal change upon P5P binding between folded and refolded GlpT WT are shown in black and open squares, respectively. Denatured protein showing lack of binding is represented by triangles. Single site binding model was found to fit best to the data giving binding constant for folded protein equal to 0.71 ± 0.11 μM and refolded protein equal to 0.84 ± 0.11 μM, error taken from the standard error of the fit. Data shown is a repeat of 3 experiments. The raw spectra were processed with CDTool with data converted into percentages, plotted and fitted applying single binding site equation using GraFit software.

CD data for folded GlpT WT in the native state fitted to the single binding site model gives a Kd of 0.71 ± 0.11 µM. Protein was then denatured and refolded by a tenfold dilution. After 10 minute incubation in the refolding buffer, P5P ligand was added, giving a total protein concentration of 0.02 $mg.ml^{-1}$. Sample was allowed to reach steady state before a CD scan was taken. Calculated binding constant from a single binding site model for refolded GlpT WT was given as 0.84 ± 0.11 µM. The denatured protein showed no change in CD signal upon ligand binding.

3.2.10 Stability of ligand bound GlpT WT

The stability of transporter GlpT WT in the native unbound state was investigated in contrast to its stability in the bound state. The protein sample was incubated for 10 minutes with various ligands prior to addition of various concentrations of urea. Unfolding curves of GlpT WT in the presence of phosphomycin only, G3P only and inorganic phosphate only are found in figure 3.21. None of the ligands above showed any change in the unfolding with urea or temperature-induced denaturation when compared to the WT. Only incubation with the P5P was found to affect the stability of the protein.

Sensitivity to the temperature denaturation of the P5P bound state of the GlpT WT was also investigated and was found to be equal to 59.5 ± 0.2 °C, observing a 10.5 °C change in Tm to GlpT WT in the absence of any ligands. Both temperature and urea denaturation curves of GlpT WT incubated with and in absence of P5P ligand can be seen in figures 3.22 and 3.23, respectively. The profile of urea denaturation changed noticeably when GlpT WT was incubated with the ligand P5P. The protein remained mostly unchanged until a concentration of 6 M urea was reached. The total percentage of CD signal loss at 8 M was similar to the urea unfolding experiment with GlpT WT without any ligands.

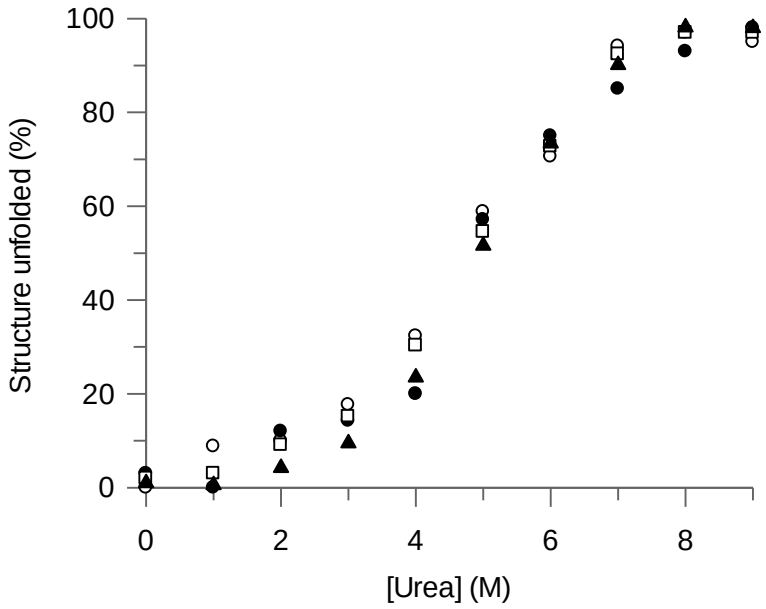


Figure 3.21 Unfolding curves of GlpT WT in the presence of phosphomycin, G3P and inorganic phosphate

GlpT WT was incubated for 10 minutes with 1 mM phosphomycin, shown as black circles, 1 mM P5P represented by open circles, 1 mM inorganic phosphate, shown as squares and control, shown as triangles, prior to 5 minute incubation with various concentrations of urea. Percentage of structure unfolded was calculated based on shift of the wavelength maximum fluorescence recorded on Fluoromax-4 fluorescence emission recordings at 295 nm excitation wavelength, in which 0 % represents protein in buffer (absence of urea) and 100 % protein unfolded at 9 M of urea. Fits of data were omitted for clarity.

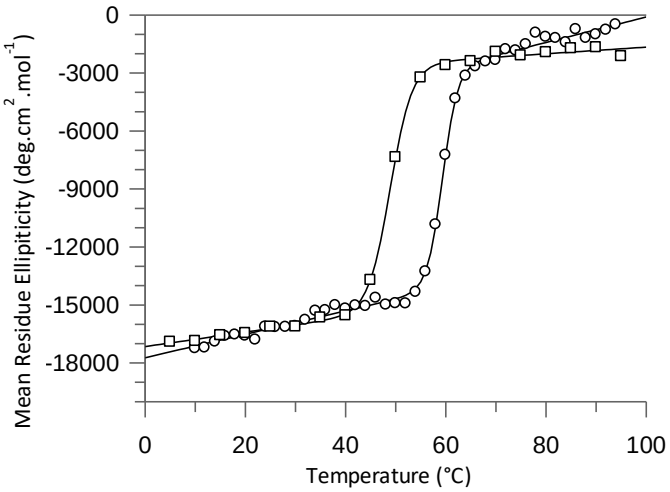


Figure 3.22 Temperature denaturation curves of GlpT WT with P5P

GlpT WT was incubated for 10 minutes at RT with 100 µM P5P, shown as open circles, before the sample was cooled down to 5° C and the temperature was increased at 2 °C intervals. The decay of CD signal at the 222 nm was used to monitor the extent of protein unfolding loss and was fit to sigmoidal curve to find the midpoint of transition upon thermal denaturation, Tm of 59.5 ± 0.2°C. Error taken from the standard error of the fit. GlpT WT denatured in the absence of ligand is shown as squares for purposes of comparison.

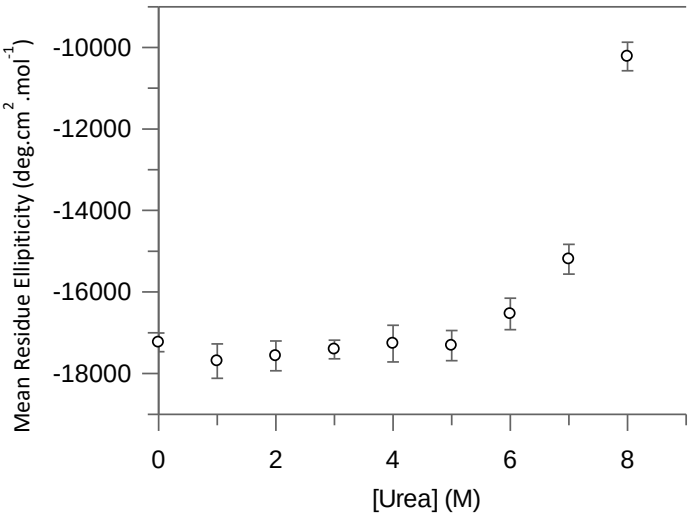


Figure 3.23 Urea denaturation curve of GlpT WT incubated with P5P

Sample of GlpT WT was incubated for 10 minutes at RT with 100 µM P5P, before the sample was denatured for 5 minutes with various concentrations of urea. The CD signal at 222 nm was used to monitor the extent of protein unfolding. No change is noted until 5 M urea in contrast to the 4.8 M midpoint of a two state denaturation mechanism of GlpT WT in absence of any ligands. Data are shown as average of 3 repeats with error taken as standard error of the mean.

3.2.11 Single tryptophan variants of GlpT

Three sites were chosen to test protein structure tolerance for tryptophan substitution. Each tryptophan mutant was analysed by CD spectrometry. The scans of single-tryptophan mutants and the folded WT are shown in figure 3.24. As described previously in chapter 2, the single tryptophan mutations were individually mutated into a tryptophan lacking gene of GlpT. The F148W mutant, located on a loop between helices 3 and 5, shows an identical spectrum to the Trp-less variant. However the M107W and V332W, located in the inter-helical bundle of domain 1 and 2, respectively, show an altered CD spectrum, implying there was an incomplete folded structure. Moreover, all the single-tryptophan mutants as well as the Trp-less variant show a smaller negative band at 208 nm, suggesting an altered structure content to WT. The mutants were also subjected to a lower yield during the purification as well as increased tendency to aggregate problems. Due to the small extinction coefficient of tryptophan, compared to the extrinsic commercial labels, a high quantity of purified protein would be necessary for time-resolved experiments. Overall, the single tryptophan mutants, were found to be non-suitable probes for unfolding, due to the difficulty in sample preparation, protein yield and altered structure. Consequently, attention was given to the rhodamine labelled variant of GlpT, described in the following sections.

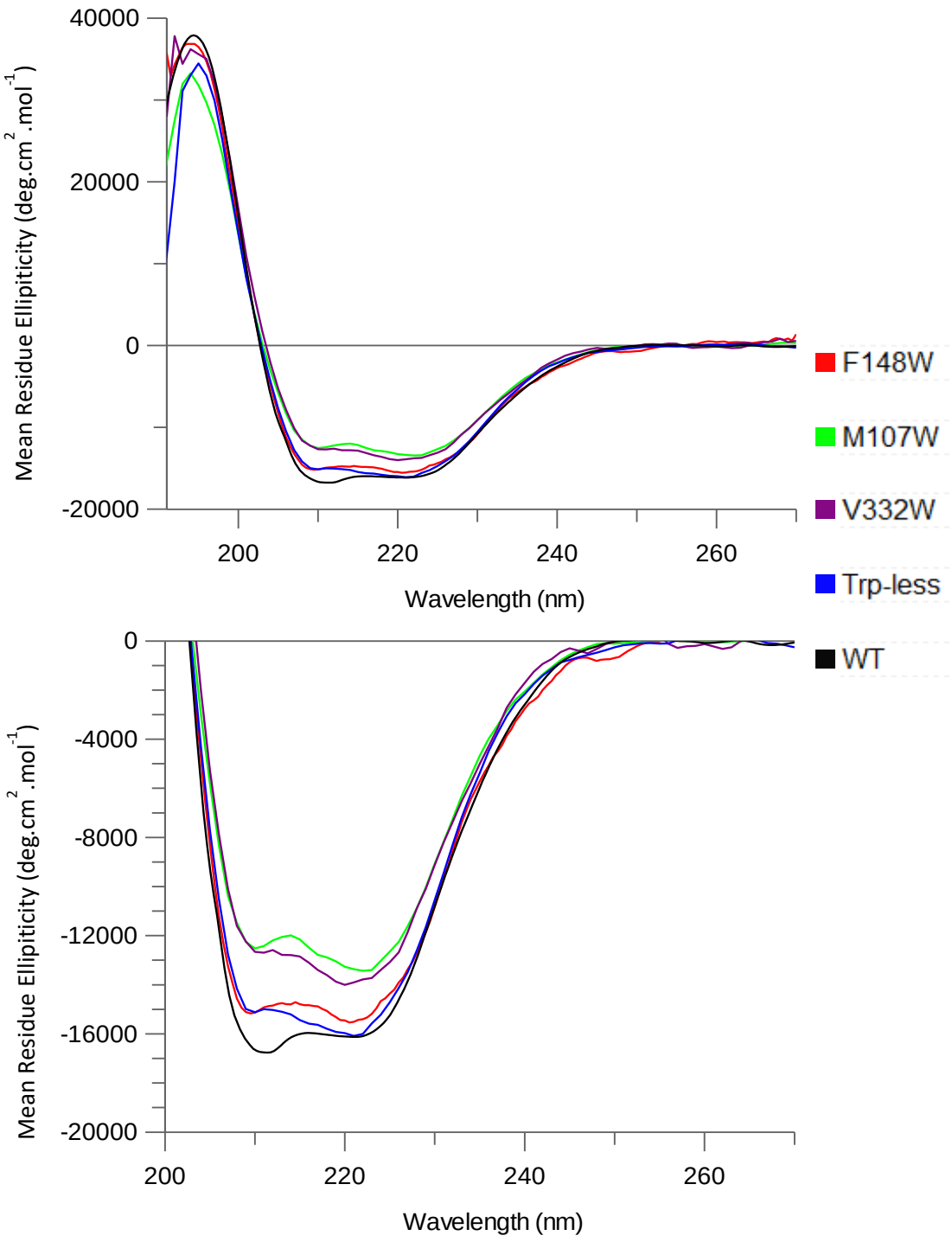


Figure 3.24 CD spectra of GlpT WT and single tryptophan variants in DDM micelles

Above, the single-tryptophan variants F148W, M107W and V332W are shown along with the Trp-less mutant and the WT for comparison. Below, the same spectra are shown expanded between 270-200 nm wavelengths to clearly show differences in the negative bands of the CD spectrum. Spectra were recorded in buffer with 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 270 nm to 180 nm. The raw spectra were first processed in CDTool then plotted and converted to MRE in GraFit software.

3.2.12 Purification of cysteine mutant GlpT and rhodamine labelling

Cysteine mutations were introduced to GlpT in order to provide sites for GlpT labelling with fluorescent probes. A ‘complete’ cysteine mutant GlpT is characterised by three mutations. Two cysteine substitutions were placed at desired label sites located at the cytoplasmic loops connecting helices 1-2 and 7-8. The remaining mutation was a serine substitution of a solvent exposed cysteine (C215S), which could interfere in the future process of rhodamine labelling. There was a marked drop in yield associated with the purification of the ‘complete’ cysteine mutant, with a significant fraction of it lost in the void volume of the gel filtration column as an aggregate. The remaining protein was collected as a monomer at an expected molecular size observed in a WT protein. The final protein elute was then spin concentrated to a stock of approximately 1 $mg.ml^{-1}$, before it was subjected to the conjugation reaction. The resulting SDS-PAGE gel of the purification procedure can be seen in figure 3.25. The far-UV spectra of the cysteine label GlpT before the process of labelling can be seen in figure 3.26. The CD spectrum of the cysteine mutant closely resembles the WT. Repeated measurements with a higher concentration of protein would improve the quality of spectra producing a more detailed comparison.

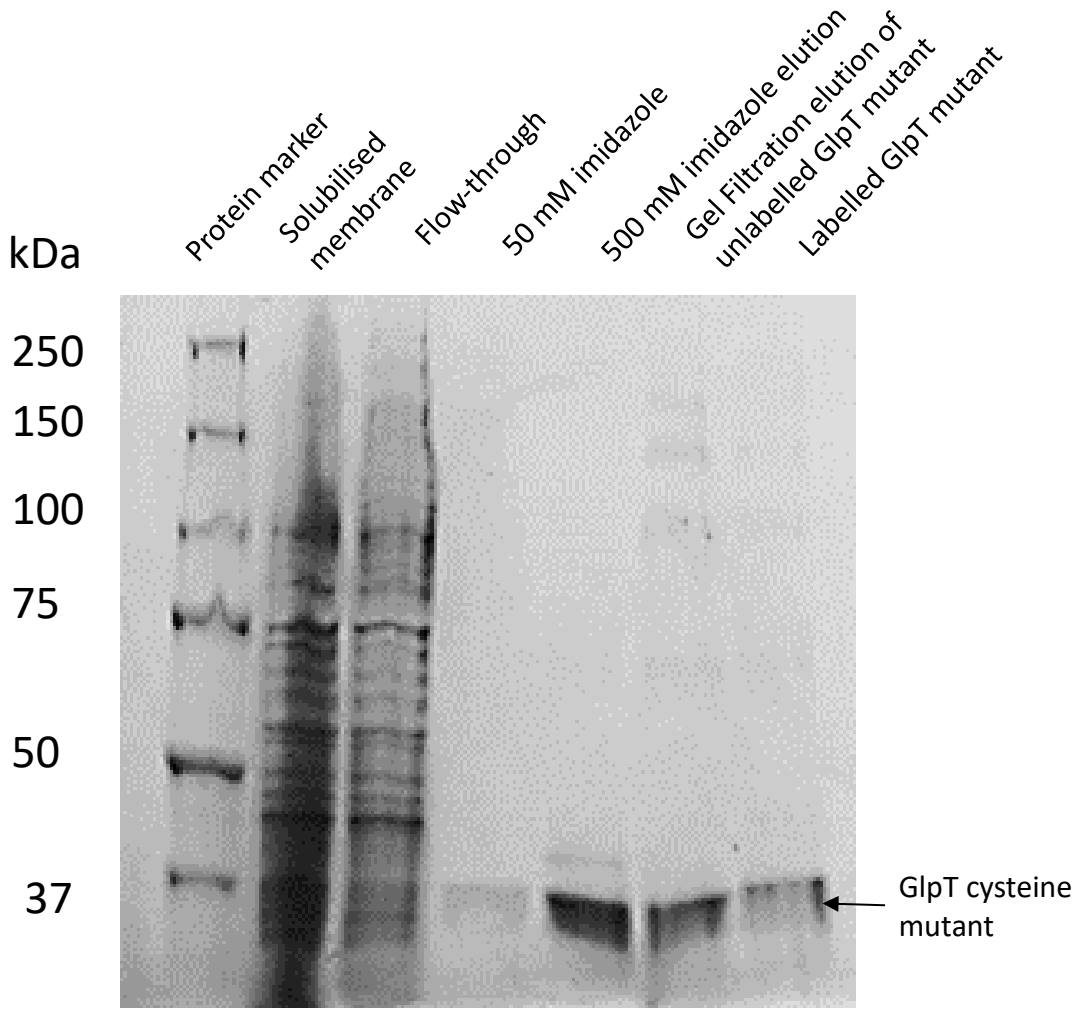


Figure 3.25 SDS-PAGE gel of rhodamine label GlpT purification visualised by coomassie blue staining

SDS-PAGE analysis was used to determine the efficiency and purity of rhodamine labelled GlpT at various stages of purification. Samples of 10 µL were taken at each stage and run on gel as follows: All-blue protein standard (corresponding molecular weight of each band labelled on the left in kDa); solubilised membrane fraction derived from the cell lysate; flow-through of an unbound fraction; 50 mM imidazole wash step; 500 mM imidazole elution collected from the Ni affinity column and spin concentrated; top gel filtration fractions pooled together and spin concentrated and finally, rhodamine labelled GlpT with the excess label removed *via* spin column. Both Ni^{2+}affinity (1 ml HisTrap HP Ni^{2+} metal ion affinity column) and gel filtration (HiLoad Superdex 200 PG gel filtration column) were carried out at a flow-rate of 1 ml\min and mounted on the ÄKTA pure chromatography system. The dye removal step was performed using a dye removal kit from Thermo Fisher.

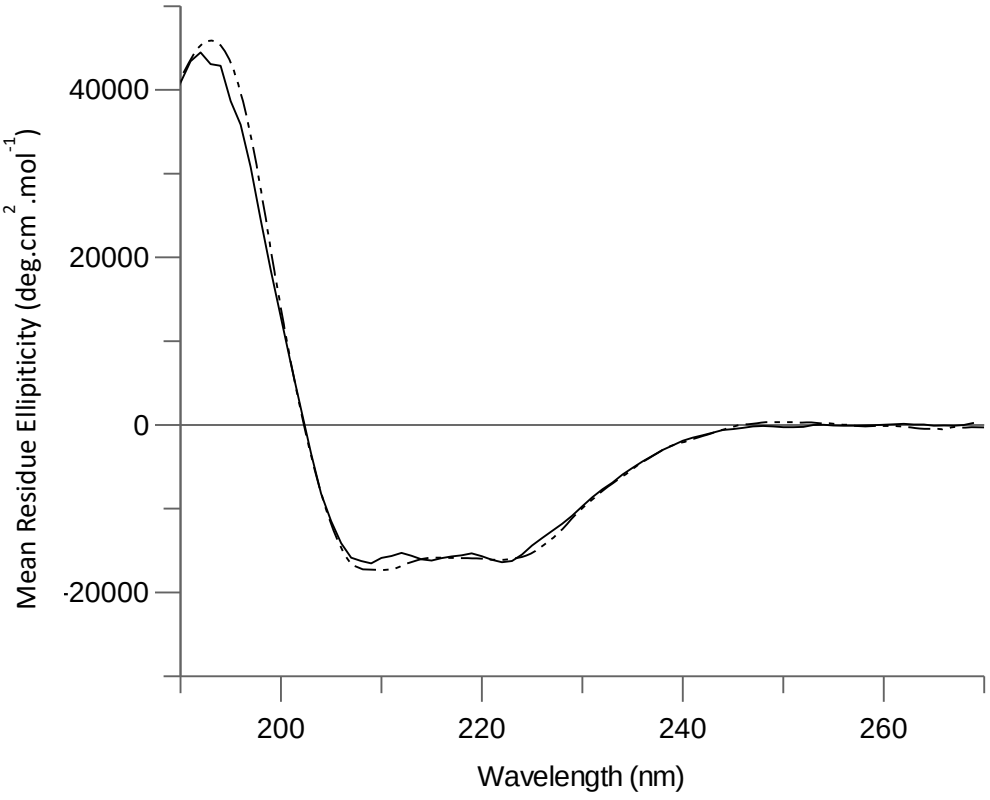


Figure 3.26 CD spectra of cysteine label GlpT and GlpT WT in DDM

The far-UV CD spectrum of cysteine mutant GlpT, marked with a solid line, and GlpT WT, represented with a dotted line was taken in native conditions to assess the similarity of the structure between the cysteine mutant and the WT. The GlpT WT and cysteine label GlpT spectra were recorded with 0.2 mm pathlength cell and 0.5 mm pathlength cell respectively, at 1 nm intervals and a 1 second acquisition time from 270 nm to 190 nm in buffer. The raw spectra were first processed in CDTool then plotted and converted to MRE in GraFit software

After the cysteine mutant GlpT was purified, it was incubated with an excess of rhodamine label. The commercial protocol comparing UV absorbance at 280 nm and 541 nm found that the recommended incubation of 2 hours labelled only 60 % of all possible label sites, yielding only a small fraction of double-labelled protein. An overnight incubation at 4°C temperature improved the efficiency by producing a yield of approximately 85 % of all possible label sites being conjugated, resulting in a better

potential for a measureable signal change in future experiments. It is possible that the DDM micelle was interfering in the labelling process, therefore a higher molar excess of the rhodamine label and extended incubation time was necessary. Care needs to be taken when estimating the labelling efficiency of the purified protein. The monomer and dimer of rhodamine give off different spectra. Therefore the resulting measurement is a collection of two different species. The efficiency of the labelling was calculated from the UV-vis spectra of protein incubated in 8 M GuHCl, as described in the methods in chapter 2. The cysteine mutant remained stable throughout the process of rhodamine labelling and subsequent dye removal process. The dye excess was removed using two dye removal columns, provided by Thermo Fisher, per labelling reaction. Approximately 10% of the protein was lost at each dye removal step. Further experiments focusing on the optimisation of labelling and purification would provide a greater final yield of the labelled protein. Moreover, additional quantification methods, for example using mass spectroscopy techniques, would identify the number of labelled residues as well as the efficiency of labelling.

3.2.13 CD measurements of rhodamine labelled GlpT in urea induced unfolding

An initial test of the newly purified cysteine mutant, GlpT C215S, was used to establish its capacity to unfold and to compare its stability to the WT. Only few CD scans of the urea unfolded sample could be collected, as the low protein yield restricted the study. The protein sample was first incubated with reducing conditions of TCEP to fully eliminate any disulphide bridges that could have formed during the experiment. The scan of the newly purified cysteine mutant GlpT shows a close resemblance to the WT, as seen previously in figure 3.26. During the urea-induced unfolding the structure of the cysteine mutant GlpT lost approximately 34 % of its structure, similar to the WT which also loses a third of its structure in 8 M urea. The full CD scan of the folded and unfolded state in DDM is shown in figure 3.27. The high noise of the scan is associated with the

low protein concentration of 0.02 mg.ml^{-1}. The profile of urea unfolding for the cysteine mutant is shown in figure 3.28. The one-time measurement of unfolding is insufficient to establish the ΔG_u^{H2O} value and the two state system. Repeats of the CD urea unfolding assay would be necessary to compare the midpoint and gradient of transition, however the results so far suggest a similar extent and profile of unfolding. The CD results are encouraging as it is beneficial that the GlpT mechanism of unfolding remained unaffected by the cysteine mutations thus enabling probing of the specific regions of the protein during the unfolding process.

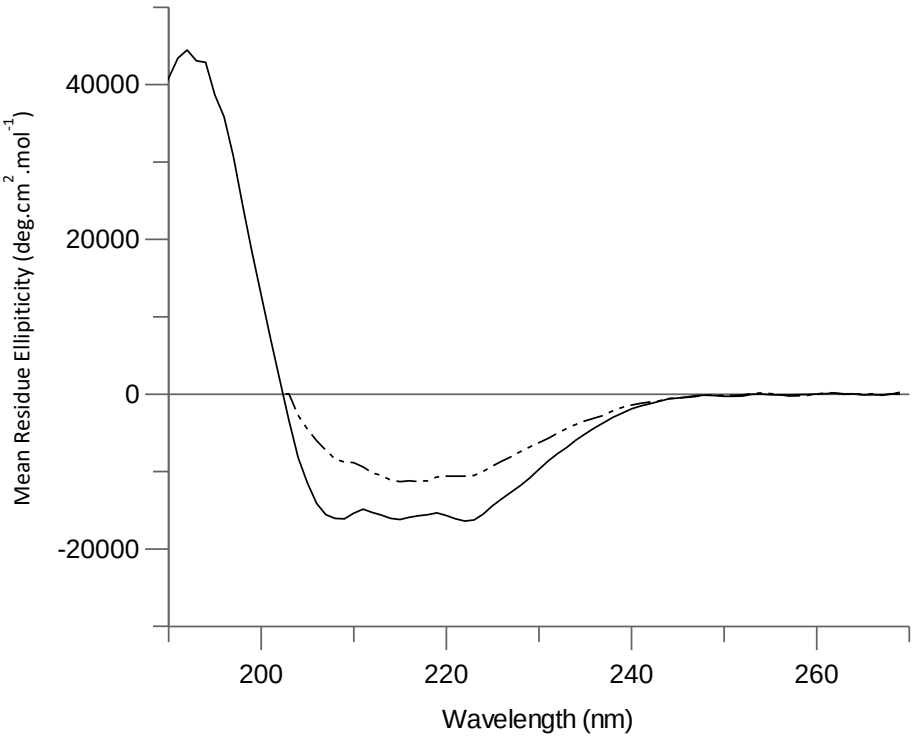


Figure 5.27 CD spectra of cysteine label GlpT and GlpT WT in DDM

Cysteine mutant GlpT was unfolded in urea for 5 minutes and the degree of unfolding monitored by 222 nm band intensity plotted against urea concentration. Spectra was recorded using 0.08 mg.ml^{-1} cysteine GlpT mutant in DDM and TCEP with a 0.5 mm pathlength cell at 1 nm intervals, a 2 second acquisition time from 270 nm to 190 nm in the buffer and from 270 nm to 205 nm in denaturant contained sample. The raw spectra were first processed in CDTool then plotted and converted to MRE in GraFit software.

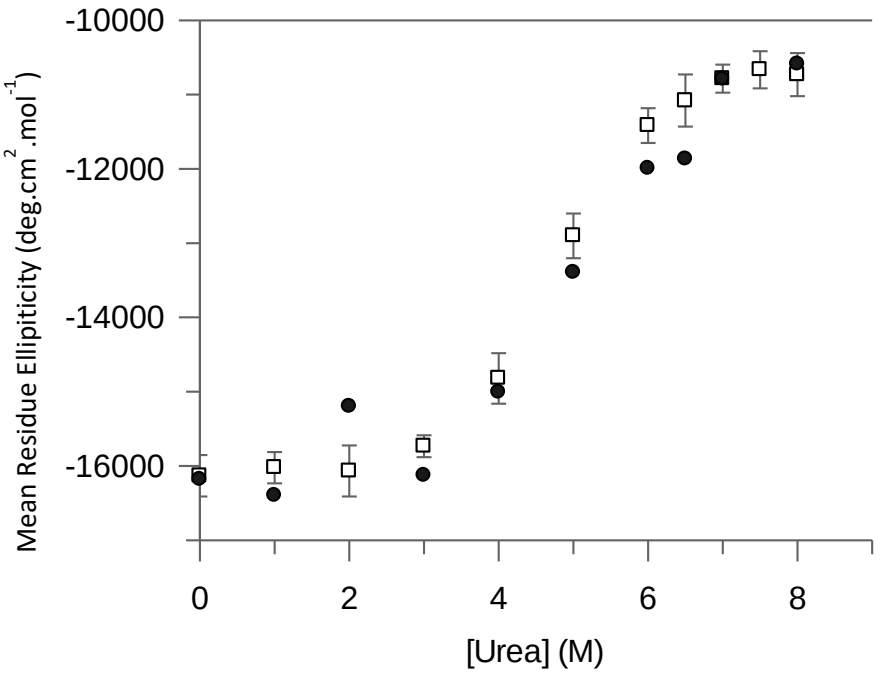


Figure 3.28 Unfolding of cysteine mutant GlpT in urea measured by CD

Cysteine mutant GlpT, shown as black dots, was unfolded in urea for 5 minutes and the degree of unfolding monitored by 222 nm band intensity plotted against urea concentration. Unfolding of the GlpT WT is shown as white squares for comparison. There was little unfolding of the cysteine mutant GlpT, shown in the first few urea concentration points, with the structure beginning to unfold after a 3 M concentration was reached, similarly to GlpT WT. The total CD signal lost between the 0 and 8 M concentration points is estimated to be around 34 %.

The stability of the cysteine GlpT mutant was further assayed *via* a thermal denaturation experiment. Similar to the CD urea unfolding assay, the sample of the protein was incubated with excess TCEP, in order to reduce any disulphide bridges that could have formed and influence the structure stability. The unfolding was monitored by the decay of a 222 nm CD signal. The CD signal was converted to MRE and plotted against the temperature, the resulting chart can be seen in figure 3.29. The calculated unfolding parameters from the sigmoidal fit show Tm of 45.6 °C ± 0.5 °C and a gradient of transition phase of 1.5 ± 0.4. Both of these are comparable with the CD measurements for WT (Tm_{WT} 48.9 ± 0.1 °C, $gradient_{WT}$ 2.2 ± 0.1) suggesting that cysteine changes did not

greatly affect the temperature stability of the protein, any differences are due to the quality of the individual purifications of the protein.

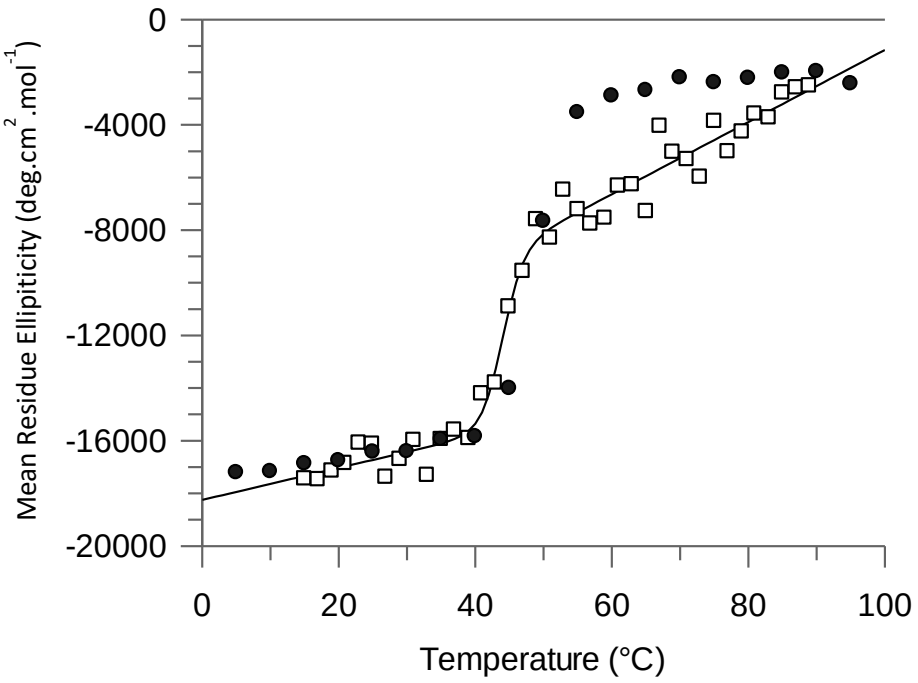


Figure 3.29 Thermal denaturation of cysteine mutant GlpT measured by CD

The cysteine mutant GlpT, shown as white squares, was incubated with TCEP prior to temperature denaturation, GlpT WT results are shown as black dots for comparison. The decay of the CD signal at 222 nm was converted to MRE and used to monitor the extent of the structure loss. The sigmoidal curve was fit to the cysteine mutant GlpT denaturation data, resulting in a Tm of 45.6 ± 0.5 °C and gradient of transition phase of 1.5 ± 0.4. The CD scans were taken at 2 °C intervals with 0.1 mg.ml^{-1} of protein sample in 1 mm pathlength cell. Data was plotted, converted and analysed using GraFit software. Errors were taken from the standard error of the fit.

3.2.14 Fluorescence measurements of rhodamine labelled GlpT unfolding

Rhodamine labelled cysteine mutant of GlpT was unfolded in urea with the progress of unfolding measured by the fluorescence response of two rhodamine labels. Both labels were placed on an endoplasmic loop located on a separate domain. The interaction between the two labels was used in monitoring the progress of unfolding seen by a fluorescence spectra recorded by 550 nm excitation in figure 3.30.

Surprisingly, denaturation by either urea or GuHCl resulted in a fluorescence loss. Approximately 21.9 % of fluorescence intensity was lost due to 9 M urea, compared to 22.5 % fluorescence lost due to GuHCl. The maximal fluorescence intensity was plotted against the denaturant concentration and can be seen in figure 3.31. The urea denaturation plot looks comparable to the two state urea unfolding plots discussed in earlier in this chapter, with a midpoint approximately near 5 M urea concentration. The current fit provides ΔG_U^{H2O} equal to 5.0 ± 1.7 kcal/mol, however the experiment would need repeating for a clearer picture of unfolding and subsequent ΔG_U^{H2O} calculation. The similarity of ΔG_U^{H2O} magnitude is encouraging, but more experiments into labelled GlpT are needed. The further fluorescence quenching recorded in the unfolding assay were possibly brought about by a further increase in proximity of the two labels in the unfolded state or contact with other quenchers on the protein side chains.

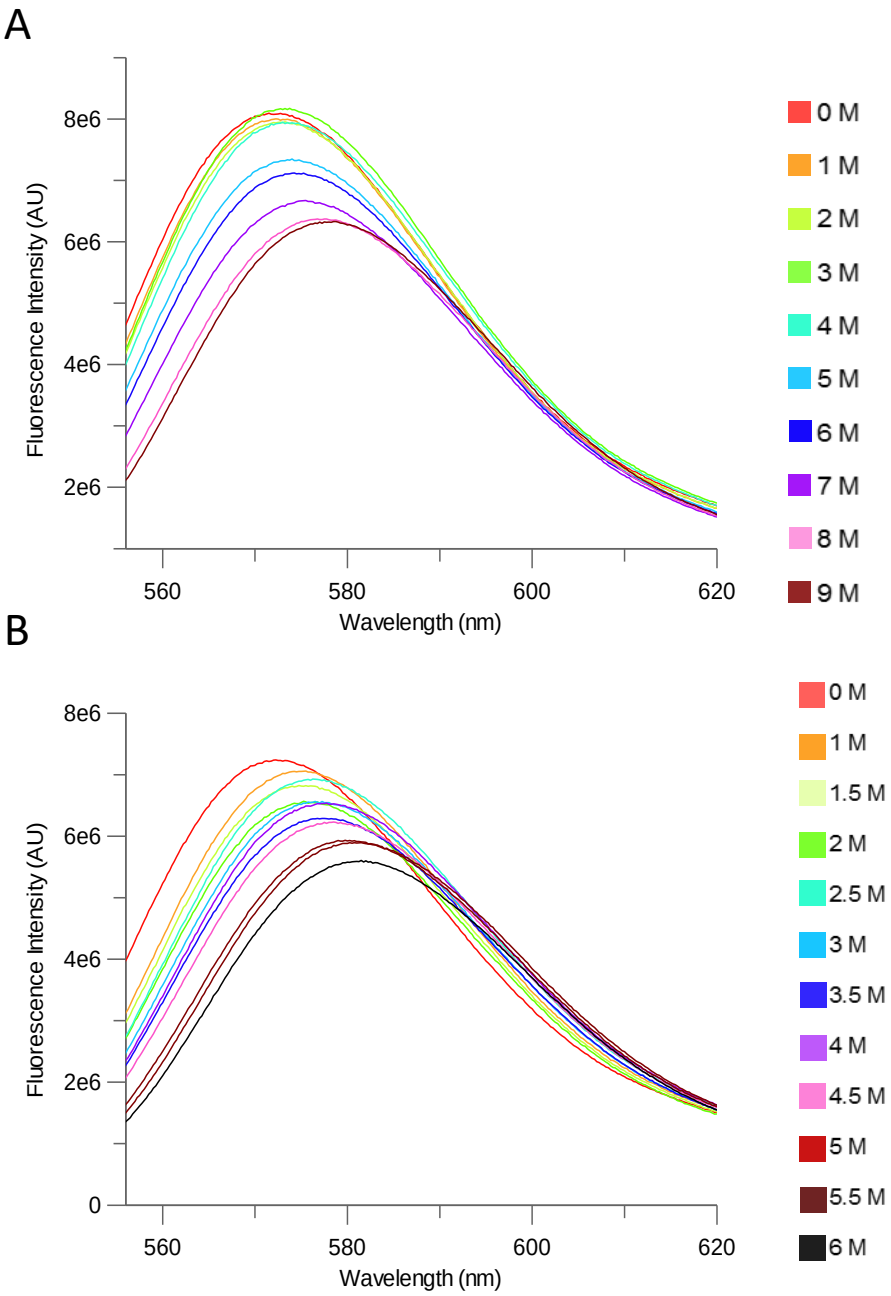


Figure 3.30 Representative fluorescence spectra of rhodamine labelled GlpT incubated wit urea and GuHCl

Representative spectra of rhodamine labelled GlpT incubated with urea and GuHCl in panels A and B, respectively. The increasing concentrations of each denaturant are shown in different colours on the right of the respective panel. A large drop of fluorescence coupled with big shifts of wavelength at maximum fluorescence can be seen in all measured spectra. The overall decrease of fluorescence due to urea and GuHCl unfolding was recorded at 22 % and 23%. All spectra were measured with 0.0025 mg.ml^{-1} of GlpT WT in a 4 mm pathlength quartz cuvette, incubated in urea or GuHCl for 5 minutes with data collected at 0.25 nm increments from between 355 and 620 nm with a slit width of 3 nm.

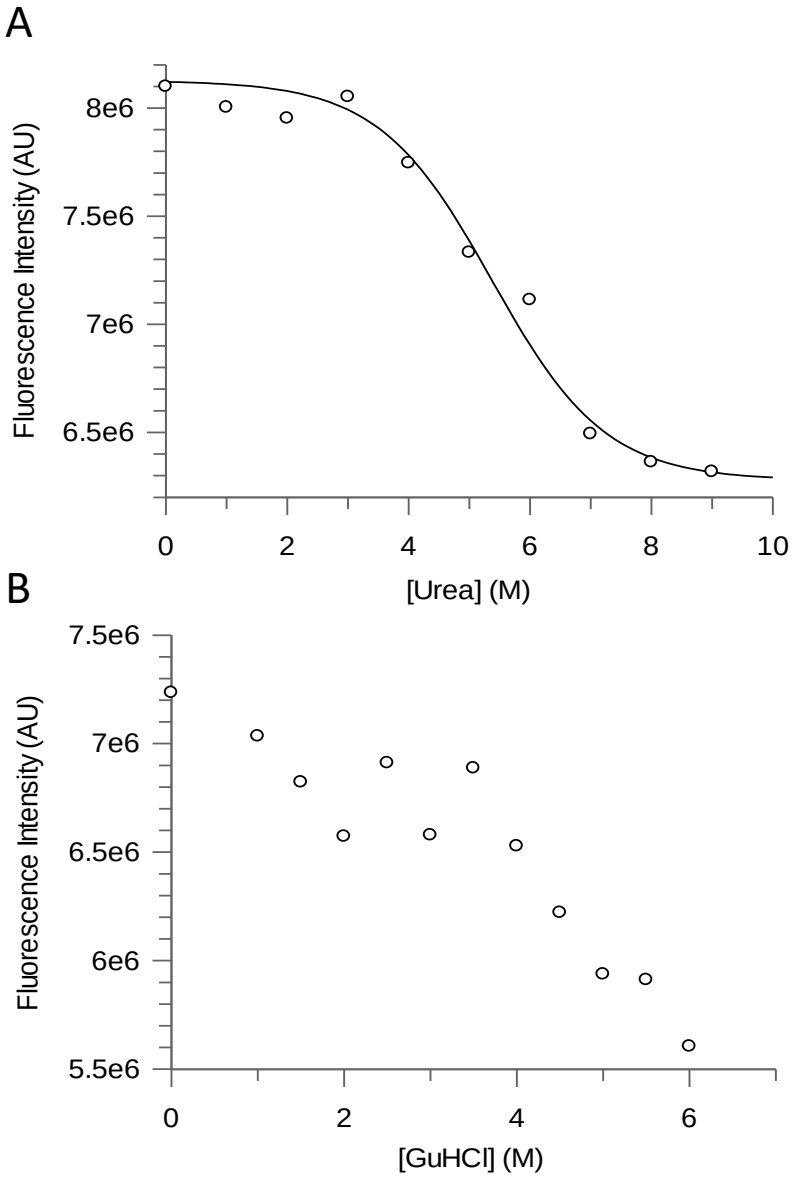


Figure 3.31 Fluorescence intensity recorded upon urea and GuHCl unfolding with rhodamine labelled GlpT

The unfolding of the rhodamine labelled GlpT with urea, in panel A, and GuHCl, shown in panel B, is accompanied with the fluorescence lost due to quenching of rhodamine labels by an increase in their proximity or due to interactions with another amino acid residue. The urea induced unfolding was fitted to sigmoidal equation with the midpoint and gradient equal to 5.4 M and 0.9, respectively. All spectra were measured with 0.0025 $mg.ml^{-1}$ of rhodamine labelled GlpT in a 4 mm pathlength quartz cuvette, incubated in urea or GuHCl for 5 minutes with data collected at 0.25 nm increments from between 550 and 620 nm with a slit width of 3 nm. Data was analysed in plotted in GraFit software.

Moreover, the label's fluorescence spectra is shifted towards a longer wavelength spectrum as the protein unfolds, as shown in figure 3.31. The wavelength shift at maximum fluorescence was analysed by fitting a normal distribution curve, as shown in chapter 3, as well as comparing the ratios of fluorescence intensity recorded at 365 and 385 nm. The resulting data was plotted against the urea concentration and is displayed in figure 3.32. The urea-induced denaturation produced a total wavelength shift of 5.7 nm compared to the 8.1 nm shift recorded for the GuHCl induced denaturation. Both denaturation graphs, show a uniform wavelength shift with difficult to distinguish pre- and post-transition baselines, making any two state fit impossible.

In the folded protein, the close proximity between the two rhodamine labels causes them to form dimer rhodamine stacks with different fluorescent properties. As the protein unfolds the ratio of rhodamine dimer to monomer population changes, affecting the fluorescence spectra recorded. Recording OD_{365}/OD_{385}, alongside the peak wavelength shift, provides an additional insight into the distribution of rhodamine monomers and dimers during the process of unfolding. The OD_{365}/OD_{385} ratio shows a slightly different profile of unfolding with urea shifting by a total value of 0.35, compared with a value of 0.41 measured for the GuHCl induced denaturation. Interestingly, there is a big change in OD_{365}/OD_{385} values recorded at the 6 M point for the urea induced denaturation and the 5 M point for the GuHCl denaturation. It is possible that at this point, the region around the rhodamine label experiences the biggest conformational change that could be represented by a transition point of the two state unfolding of GlpT. If that reasoning is correct, then it is possible that the periplasmic side of GlpT, where the two rhodamine labels are present, experiences unfolding at a later stage than the rest of the protein.

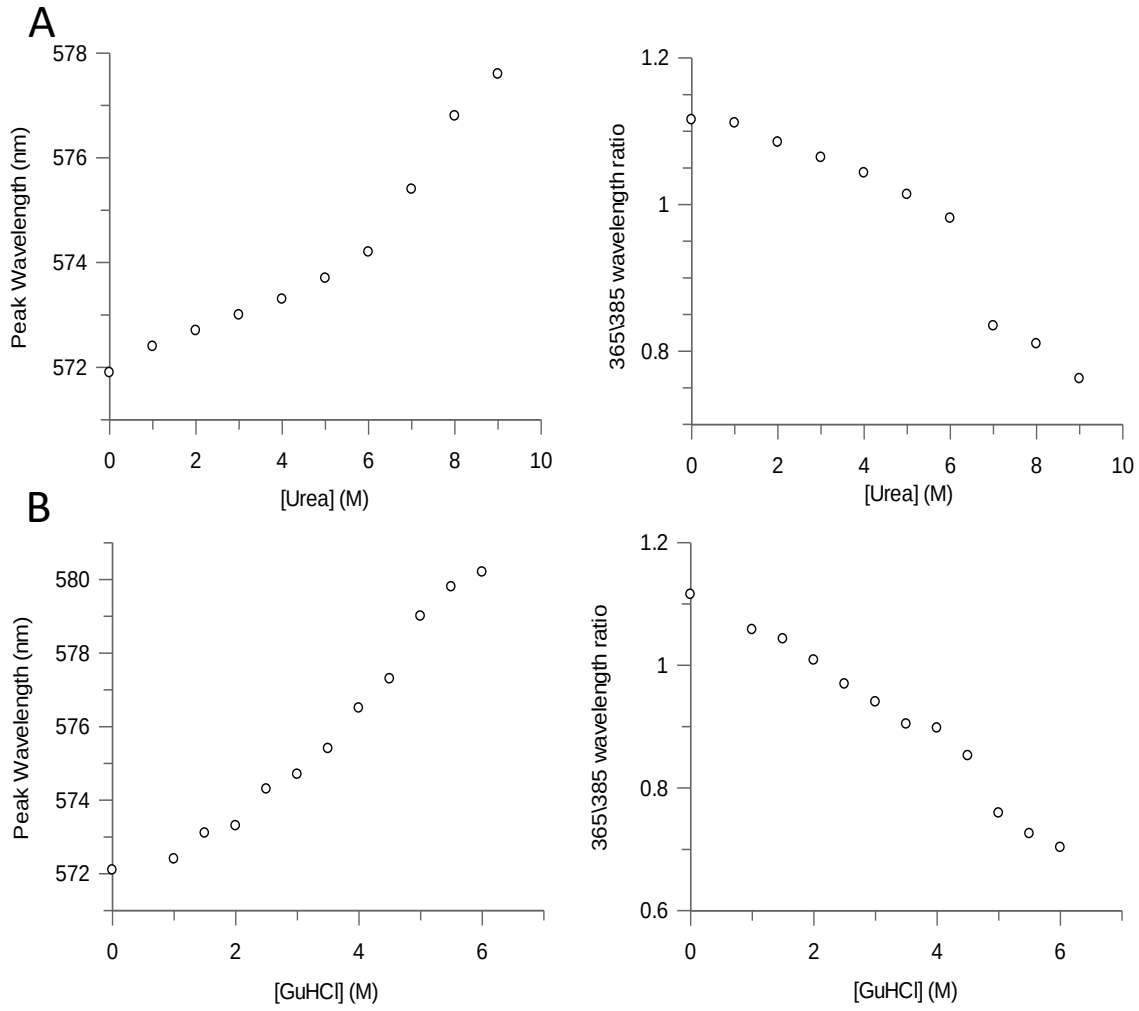


Figure 3.32 Wavelength shift recorded upon urea and GuHCl unfolding with rhodamine labelled GlpT

The protein unfolding with urea, panel A, and GuHCl, shown in panel B, is coupled with the wavelength shift used to monitor the degree of unfolding. The wavelength shift of 5.7 nm was recorded for the urea induced denaturation, compared with 8.1 nm with the GuHCl. The ratio of fluorescence intensities at 365 and 385 nm. This was used as an additional measure of the fluorescence peak shift, due to a rhodamine dimer formation, the ratio analysis can be seen on the right with their respective peak wavelength shift graph. The OD_{365}/OD_{385} ratio shows a slightly different profile of unfolding with the urea shifting by a total value of 0.35, compared to 0.41 value measured for the GuHCl induced denaturation. All spectra were measured with 0.0025 $mg.ml^{-1}$ of GlpT WT in a 4 mm pathlength quartz cuvette. Data was analysed and plotted in GraFit software.

3.2.15 Fluorescence measurements of a refolding reaction of rhodamine labelled GlpT

The refolding capacity from the urea denatured state of the rhodamine labelled GlpT was assayed by fluorescence measured at 550 nm excitation wavelength. The fluorescence spectra of the folded and refolded sample with the unfolded spectra for comparison can be seen in figure 3.33. Approximately 95 % of total fluorescence is recovered. The refolding is also coupled with the return of the native wavelength at maximum fluorescence intensity, implying rhodamine labels return to their original environment in the folded state. Similarly to GlpT WT, no refolding was observed from the GuHCl denatured state. The progress of refolding at various urea concentrations was followed by the wavelength shift and can be seen in figure 3.34. The refolding and unfolding profiles show close similarities, implying a similar folding and refolding pathway. The fluorescence refolding experiments would need repeating for a more complete picture of refolding, however data so far is encouraging as it suggests minimal impact of the labelling and cysteine mutation on protein behaviour seen in earlier in this chapter. Further experiments focusing on optimisation of labelling and protein purification would solve the issue. Similarly, the ligand binding assay would have to be performed to assess the effect of mutations on the ligand binding affinity as well as the return of the tertiary structure around the binding site.

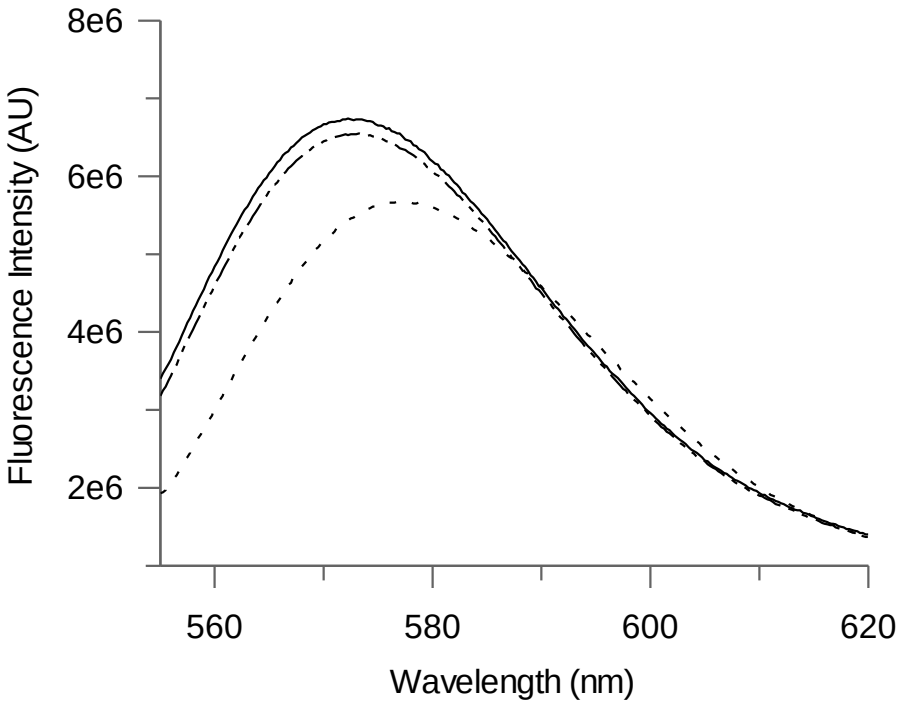


Figure 3.33 Representative fluorescence spectra of the folded, refolded and unfolded rhodamine labelled GlpT

The rhodamine labelled GlpT was refolded by a ten-fold dilution of the unfolded GlpT sample incubated with 8 M urea. The overlay of refolded and folded samples, shown here in solid and dashed lines respectively, represents the almost complete return of original fluorescence. The protein was unfolded for 5 minutes in 8 M urea at room temperature giving the spectra seen by a dotted line. The refolded sample was unfolded in the same manner, followed by a ten-fold dilution into the buffer and incubated for 10 minutes at room temperature before taking the measurement. The scans of all samples were taken with 0.0025 $mg.ml^{-1}$ of GlpT WT in a 4 mm pathlength quartz cuvette, with data collected at 0.25 nm increments from between 355 and 620 nm with a slit width of 3 nm.

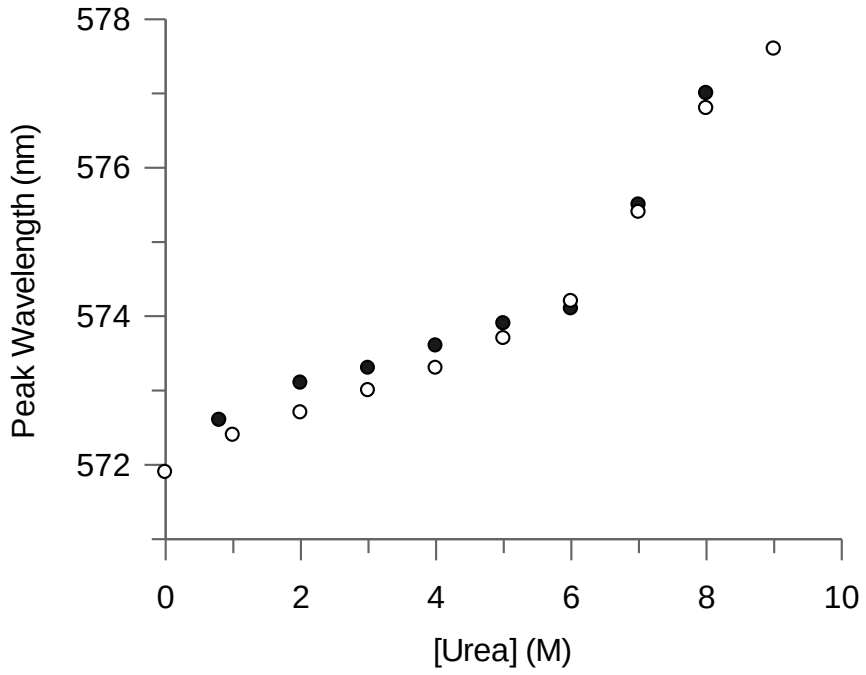


Figure 3.34 Comparison of rhodamine labelled GlpT unfolding and refolding profiles of wavelengths shifts in the urea denaturation experiments

The progress of refolding was monitored at each urea concentration by estimating the peak wavelength using a normal distribution fit. The refolding data, shown by black circles, is compared to the unfolding data displayed earlier in figure 5.11 and represented here by white circles. Both profiles overlay closely with each other with a small difference present at 0.8 M point, resulting in an almost complete return of the native peak wavelength. The data was analysed and plotted using GraFit software.

3.3 Discussion

3.3.1 Stability of GlpT WT to denaturants

The stability of GlpT WT was investigated using chemical denaturants such as urea and GuHCl. Approximately 33 % of protein CD signal was lost upon denaturation with urea. Alternatively, the stronger denaturant GuHCl was found to unfold up to 45 % of the protein. GlpT WT unfolding was also investigated by fluorescence emission experiments excited at both 280 and 295 nm. These experiments provided different ΔG_U^{H2O} values of 4.5 and 4.8 kcal.mol^{-1}, respectively. The difference between ΔG_U^{H2O} derived values is difficult to interpret, as the abundance and wide-spread positons of aromatic residues across the GlpT structure, complicate the interpretation. Changes affecting one tryptophan can be offset by interactions with other aromatic residues. Any change to fluorescence spectra can therefore be misinterpreted as little unfolding happening. Therefore, fluorescence experiment results are mainly used as a complementary picture of GlpT unfolding. The discrepancies between different fluorescence experiments, from using different excitation wavelengths, suggest that a CD technique is better in measuring ΔG differences in the chapter 4, as it represents a global change in the protein state. Although GuHCl unfolds the protein to a much larger extent it is more difficult to assess the pre-transition baseline of unfolding. Alternatively, the unfolding reaction using urea shows clear pre- and post- transition baselines, as well as reversibility, which in turn makes the urea unfolding assay appropriate to calculate ΔG_U^{H2O} values. The following chapter 4's analysis of alanine scan is therefore based on urea denaturations.

Superimposed unfolding curves and similar fitting parameters from fluorescence and CD experiments show that the loss of structure is closely followed by an increase in solvent exposure indicated by red shift of the wavelength maximum, as well as fluorescence intensity quenching. Furthermore, during the process of unfolding, GlpT losses a third of helicity, combined with only a small wavelength shift. These outcomes suggest that the GlpT unfolding process results primarily in disrupted helix interactions, while most

aromatic residues stay protected at the expense of loss of structure. It is likely that the extent of helix structure unfolding is mostly limited to the ends of helices and possibly to disrupting coiled coil helix interactions of two domains of GlpT. The comparison between CD and fluorescent measurements hints at a non-uniform GlpT unfolding, where different areas of protein unfold to a different extent. Additional studies using fluorescence probes, sensitive to changes of environment would help to obtain more localised information, which would then help to determine the extent of unfolding in individual regions of the protein.

3.3.2 Temperature denaturation

Thermal denaturation was performed to assay the stability of the GlpT WT, as well as to compare it to the temperature stability of other MFS proteins. The assay is also used to measure the differences in stability of GlpT's alanine mutants in chapter 4. The process of protein unfolding due to increasing temperature was monitored by both fluorescence and CD. Both experiments were fitted to the sigmoidal equation to calculate the melting point, Tm. A slight difference was observed in the calculated values for Tm, with the fluorescence recorded Tm value being a few degrees larger. The difference could be attributed to the error in protein concentration, which was found to vary with the measured Tm. Therefore, to provide comparison, all protein temperature denaturation experiments were performed at the 0.1 $mg.ml^{-1}$ protein concentration. Bigger protein concentrations increase the risk of protein aggregation and therefore the resistance to temperature denaturation. However, lower concentrations decrease the CD signal, leading to an increased error of the measurement.

The measurements of fluorescence at a range of temperatures showed a great dependence of fluorescence on temperature, with a steep pre- and post-transition baseline recorded for the fluorescence intensity. These results suggest great flexibility of GlpT structure to respond to changes in environment, exposing the aromatic amino acids to the surrounding solvent. While fluorescence shows a sharp baseline, structure

changes are less obvious at the pre-transition stage. The results suggest that as temperature increases and the protein unfolds, first the tertiary contacts are lost with most of structure loss occurring at the most solvent exposed areas of the helices. Followed by the loops and then in the final stage, the alpha helices unfold. However, this needs be further investigated using methods that provide more specific measurements on stability of individual regions of the protein. The effect of individual alanine substitutions on the temperature stability of GlpT is later discussed in chapter 4.

3.3.3 Reversibility of the GlpT WT unfolding

Unfolding of GlpT WT from urea induced denaturation was found to be reversible by the complete return of the CD signal measured at the 222 nm. Furthermore, the full CD spectra of the refolded sample was found to overlay with the scan of the protein before denaturation. These results are indicative of the native secondary structure return upon removal of the denaturant. Once the concentration of urea was reduced to a lower range, the secondary structure contacts were re-established to the same extent, as shown by the refolding curve comparison of the secondary structure content between the unfolding and refolding curves at each different urea concentrations. The overlay and close resemblance of both unfolding and refolding curves is a good indication that forward and reverse changes belong to the same reaction. Further evidence is provided by comparison of ΔG values from folding and unfolding curves. If both curves belong to the same reaction, then the ΔG values should be similar. This is indeed the case as the folding and unfolding curves of GlpT WT overlay against each other, with ΔG_u^{H2O} and ΔG_F^{H2O} both found in close proximity to the 4.6 kcal/mol value.

To further assay GlpT reversibility, information on the return of tertiary contacts is also required. The results from P5P binding assay show closely related binding affinities measured in the sample before denaturation and after the refolding reaction. This suggests that the GlpT is able to regain its secondary structure as well as tertiary conformation, as the amino acid residues responsible for binding have likely returned to

their native positions before denaturation, thus resulting in a similar affinity to substrate binding.

3.3.4 P5P binding assay

A ligand binding assay was primarily developed to assay the reversibility of GlpT unfolding. The effect of ligand binding on protein stability as well as the effect of alanine mutations on the GlpT functionality were also investigated.

The mechanism of GlpT ligand binding was first described by Law (117, 119). However, the previously published protocol was not successfully reproduced. The assay suffered problems with binding not causing significant changes of fluorescence when substrate was introduced. Furthermore, photobleaching was found to cause too much interference during the titration experiments, making substrate induced fluorescence quenching difficult to analyse (173). The assay was modified to follow the extent of ligand binding by measuring changes of CD signal upon addition of substrate. Both the G3P and Pi, previously tested in fluorescence assay, did not show significant changes in the CD spectra. Out of the remaining substrates tested, the P5P based assay was found to be the most reliable and robust method of calculating ligand binding affinity and hence is further used in chapter 4's estimation of alanine mutants' reversibility.

The results from P5P binding assay showed the binding constant, Kd, to be 0.84 μM. The previously discussed refolding experiment showed the return of the native binding, suggesting the return of tertiary structure and function upon reduction of denaturant concentration. The single binding fit was most accurate for both experiments, agreeing with the theory that P5P acts as inhibitor of G3P transport by binding near the G3P binding pocket. The change in CD signal upon P5P binding also suggests that the P5P mediated inhibition of G3P transport could be associated with conformational changes, however this needs further investigation.

The stability of P5P bound state was also investigated. The P5P bound protein showed both increased temperature and urea stability. The dramatic change of urea resistance could suggest that the P5P-bound state has less solvent exposed pockets, resulting in urea denaturation sites being less accessible to the denaturant. Surprisingly, P5P was the only ligand to invoke such change in stability, suggesting a different binding mechanism compared to other GlpT substrates: G3P, Pi and phosphomycin. Combined with the increased temperature resistance, it is possible that P5P bound state is less flexible, prohibiting interaction of other ligands. The binding of P5P uncovered interesting dynamics about GlpT and provides some information on how this interaction affects inhibition of G3P/Pi transport. These conclusions need to be further investigated before any clear mechanism is uncovered. It is equally possible that the DDM micelle has also affected these interactions by blocking the G3P-Pi binding site, but not P5P nor phosphomycin site. A series of silent mutations along the binding pocket could determine the similarities between P5P and G3P binding sites. Two of these mutations Y38A and Y266A are discussed in chapter 4.

3.3.5 Stability of cysteine mutant measured by CD

The cysteine mutant was found to unfold to a similar degree in urea compared to the WT variant, losing approximately a third of the structure in 8 M urea. Due to the necessity of sufficient protein concentration to gain accurate measurements, further protein preparations are required to ascertain ΔGu^{H2O} value. It would be useful to be able to compare this value to the ΔGu^{H2O} of the WT, in order to fully understand the effect of mutations on the stability of the protein. A suitable cysteine mutant would have similar stability to the WT with the additional advantage of having an accessible labelling site.

The loss of secondary structure due to temperature was also measured *via* CD spectrometry. No significant differences were found in the Tm measured for the WT and cysteine mutant. The reduction of –SH groups on cysteines eliminated possible formation of salt bridges and associated rise in temperature stability.

3.3.6 Fluorescence unfolding and refolding of cysteine mutant

The changes in the fluorescence spectra in one area of the protein compared to the data from other regions would ultimately be used to gain information on stability of specific regions of the protein. Currently, the data is available from only the periplasmic site of the protein in the vicinity of helices 1 and 7.

The data collected so far detects a big shift of the wavelength of the band maximum after the 6 M urea concentration point (3.4 nm shift) and 4 M GuHCl point (3.7 nm shift), a result that could mean slowed unfolding of the periplasmic region. The shift could be caused by an interaction with the polar residue, such as lysine. Alternatively, the shift could be caused by an exposure to the surrounding solvent, a result that is indicative of the protein unfolding. Substitution mutation of the neighbouring lysine would eliminate its involvement in the wavelength shift, confirming the nature of the fluorescence spectra shift.

A possible explanation of the delayed unfolding of the periplasmic region is provided by the crystal structure of GlpT in the open state. Without a ligand, the periplasmic side of the membrane remains closed to the transport of substrates by concealing the binding cavity. A previous single tryptophan study of LacY found that when a mutation blocking the binding cavity was introduced the protein gained stability. This is due to the limited accessibility of denaturant to the protein core. Only when the denaturant concentration exceeds a certain threshold, the denaturant is able to overcome this barrier leading to complete protein unfolding.

A previous stability experiment with P5P-bound GlpT has shown that ligand binding accompanied with the bound-state conformational change can drastically alter the protein resistance to denaturation. Interestingly the measurements of fluorescence intensity do not seem to confirm the late unfolding step of the periplasmic region. It is possible that other factors affecting the fluorescence intensity masked the unfolding step,

such as affinity of two rhodamine labels to each other. Furthermore, the measurements of fluorescence intensity are more susceptible to error from the wrong protein concentration. Repeats of the experiment would be necessary to fully discuss the involvement of the periplasmic region in the unfolding pathway of GlpT. Also, other connections than cysteine links can be made. During incubation with a high concentration of TMR, it is possible to bind TMR to, for example lysine (169). Therefore, care needs to be taken when estimating the labelling efficiency of the purified protein.

Further rhodamine quenching could be explained by an increased proximity of rhodamine labels and\or interactions with surrounding amino acids (173). The summary of possible interactions is displayed in figure 3.35. Previous studies of extrinsic labels have found that amino acids, such as: tryptophan, tyrosine, histidine, methionine and aspartate were responsible for some rhodamine quenching effects. Out of all natural amino acids tryptophan was identified as a main quencher of rhodamine (174).

An additional test between rhodamine and denaturant interaction should also be performed. The previous study on associations between titin molecules (175) observed a decrease in fluorescence upon increase of denaturant concentration. They have concluded that a further decrease in fluorescence intensity was due to the direct interaction of chemical denaturants (GuHCl, urea and SDS) with the rhodamine markers, through an unspecified mechanism of quenching; either collisional or through dark-complex formations. Although this explanation is likely and needs further experimentation with GlpT constructs, the decrease in florescence measured did not follow the linear relationship with the concentration of the denaturant. Further analysis would likely answer this proposition.

The refolding from the urea-induced state was shown for the cysteine mutant of GlpT by the return of a native fluorescence spectra of the rhodamine label. Both the peak wavelength and the fluorescence intensity return to their original values. A small discrepancy is noted, that could've been caused by a drop in efficacy of refolding. Since the results come from a single experiment, repeats are necessary to define whether or not the refolding profile matches the unfolding. It is unlikely that the cysteine mutations caused a change in the usually robust refolding protein such as GlpT, however it is necessary to confirm the results with the return of the secondary structure as measured by the CD as well as ligand binding assay to probe the return of the tertiary structure.

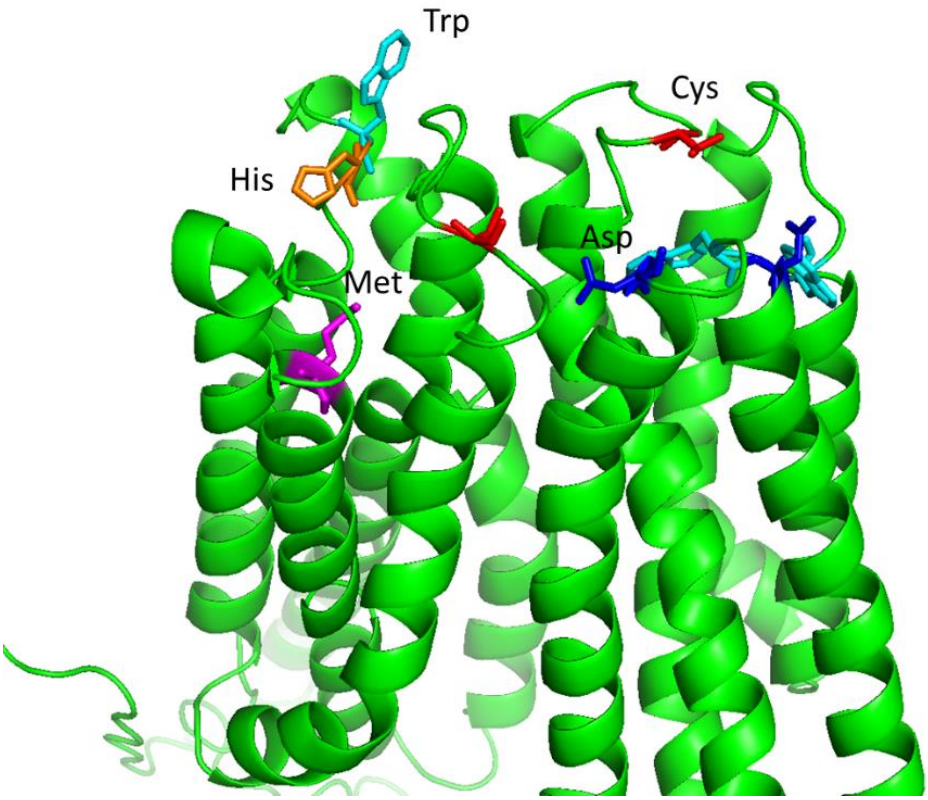


Figure 3.35 Crystal structure of GlpT's periplasmic region in the context of cysteine mutations

The mutated loops are located between helices 1-2 and 7-8. The cysteine mutations are shown in red, the nearby charged residues are coloured blue for aspartate, polar residues are coloured orange for histidine, purple for methionine and cyan for tryptophan. Structure was rendered in PyMOL software.

Overall. All results from the rhodamine labelled and cysteine mutant experiments of GlpT are incomplete and need further work to be done before definitive conclusions can be made. However, results presented in this work show a potential for such studies as well as possible methodology that could be adopted.

3.3.7 Comparison with the other MFS transporters

Only a handful of investigations are dedicated to studies of multi-subunit membrane protein folding. When comparing results from these experiments, it is important to compare proteins that were studied under similar conditions. The difficulty of membrane protein *in vitro* work has led to a development of a number of different solubilisation systems. Furthermore, a range of different denaturants is often used with membrane proteins as their common feature resistance to denaturation (176). The ability to compare GlpT's behaviour during unfolding is therefore limited to a handful of other membrane proteins that were also studied under similar detergent\denaturant conditions.

The MFS transporters, LacY, GalP and GlpT were solubilised in DDM and all show a loss of approximately third of the total CD signal when denatured with 8 M of urea. In all three proteins, the pre-transition unfolding phase looks similar. There is little change to either fluorescence or CD signal, when proteins were incubated with low concentrations of urea. Only when proteins were exposed to concentrations of 3 M of urea, did they start to unfold with a midpoint of unfolding curve of around 4 M. This emphasises alpha helical membrane protein's shared feature to resist denaturation.

GlpT, as well as other members of the MFS, LacY, GalP and XylE when unfolded with urea display urea denaturation reaction that exists in a reversible equilibrium. The return of native structure before denaturation was shown for all these protein *via* far-UV CD in conjunction with the return of substrate binding. Thus, MFS members so far have been shown to be good candidates for folding studies due to their robust refolding ability. However, there have only been a few successful attempts in obtaining full kinetic data

for microscopically reversible membrane proteins in dynamic equilibrium (59, 177). Performing kinetic analysis on a large two domain protein, such as GlpT, is complicated and although it requires more experiments, it might be possible to obtain more complete kinetic data with help of single fluorescent tags.

Unfolding experiments for GalP and LacY have shown the ΔG_U^{H2O} of 2.5 kcal.mol^{-1} and 2.2 kcal.mol^{-1}, respectively. The size of the protein was found to be related to the degree of protein stability and resulting ΔG_U^{H2O} value (178). GlpT is an example of a larger protein, with molecular weight of 52 kDa compared to 42 kDa LacY, with a measured ΔG_U^{H2O} equal to 4.3 kcal.mol$^{-1.}$ Interestingly, GalP's molecular mass is estimated to be equal to 50 kDa, similar to GlpT's 52 kDa, however ΔG_U^{H2O} of GalP is only 2.5 kcal.mol^{-1}, compared to GlpT's 4.3 kcal.mol^{-1}. Regardless of size or structure similarity between these 3 MFS transporters, GlpT is an outlier in regards to its calculated free energy of unfolding. It is likely that ΔG_U^{H2O} is linked with the exposed surface of the protein, rather than the protein's size. However, taking into account the effects of detergent micelle, it is difficult to estimate the actual denaturant exposed surface of the protein.

Moreover, GlpT has higher energy of unfolding per residue of 0.10 kcal.mol^{-1}.residue^{-1}, similar to DGK (0.13 kcal.mol^{-1}, 153) and bR (0.08 kcal.mol^{-1}, 73), unlike the GalP and the LacY with a value of 0.05 kcal.mol^{-1}.residue^{-1}. According to the rocker switch mechanism of MFS transport, transporters undergo significant conformational changes in order to expose large hydrophilic binding sites to their substrates. The inherent MFS flexibility and dynamic structure, needed to facilitate substrate transport, is thought to be linked to a greater exposure to denaturants and therefore result in lower unfolding per residue energy and ΔG_U^{H2O} values for GalP and LacY. However, compared to GalP and LacY, GlpT has a higher unfolding per residue and ΔG_U^{H2O} value. Unlike GalP and LacY's symporter transport cycle, GlpT is an antiporter. Different mechanism of transport and therefore accessibility to denaturant, could provide an explanation for a higher reported stability, and deviation from low energy of unfolding per residue reported for LacY and

GalP. This suggests that, regardless of common fold between MFS transporters, there are differences in protein's chemical sensitivity. Furthermore, protein's mechanism of transport may also be linked to the protein stability.

3.4 Summary of chapter

- GlpT WT two state analysis of the folding reaction measured by CD reports ΔG_U^{H2O} of 4.3 kcal.mol^{-1}.
- Fluorescence measurements of ΔG_U^{H2O} report values of 4.5 and 4.8 kcal.mol^{-1} for measurements taken at 280 and 295 nm excitation wavelength, respectively.
- The binding of P5P can be measured by a ligand induced change, measured by CD spectroscopy, reports Kd of 0.71 µM.
- Urea induced unfolding is reversible and is shown by return of secondary structure and P5P binding.
- Unfolding of cysteine mutant, measured by CD spectroscopy, shows characteristics of a two-state behaviour, as seen in WT.
- Urea-induced unfolding of cysteine mutant is reversible.
- Unfolding of rhodamine labelled GlpT causes a decrease of fluorescence due to further quenching of the labels.

Chapter 4

Alanine substitutions in GlpT

4.1 Introduction

4.1.1 Alanine substitutions of GlpT

In the chapter 3, GlpT's unfolding and refolding experiments were recorded by both fluorescence and CD spectroscopies. Furthermore, the free energy of GlpT WT unfolding was established, along with the ligand binding assay to demonstrate GlpT function. Both the CD monitored refolding assay and ligand binding assay have shown GlpT to follow the reversible equilibrium reaction when urea is used as a denaturant. The methodology established in chapter 3 is used in the following alanine scan experiments.

As mentioned previously, studies performed on LacY, suggest that certain protein regions show increased stability and may also pose a role in the process of integration with the membrane (125). The aim of this study is to examine whether the stability of the GlpT structure could also be localized to the first helix of the first domain, by identifying whether or not the first helix shares the same stability as structurally similar the first helix of the second domain, helix 7. If GlpT shows similar stability tendency, like LacY, then it could indicate a folding trend in other MFS transporters. Each alanine substitution on helix 1 will have its corresponding mutation located on the opposite site on helix 7, the first helix of the second domain. This second mutation will be selected based on a similar amino acid substitution change as well as a similar electrostatic environment to its respective helix 1 mutation.

4.1.2 Comparing helices 1 and 7

The helices 1 and 7 of GlpT are the first helices of the first and second domain, respectively; figure 4.1 shows the structure of GlpT with highlighted helices 1 and 7. Based on the whole residue octanol scale hydrophobicity plots, showing preference of certain amino acids to hydrophobic parts of the membrane, and resolved X-ray crystal structure, Huang (97) estimated the length of the GlpT's helix 1 to be 34 amino acids long compared to helix 7 which is made up of 37 amino acids. Both helices 1 and 7 are

located in similar central positions within the protein structure, lining the binding cavity. They both exceed the width of the bilayer and are hence tilted at a similar angle of 20° from the membrane interface, in order to hide the hydrophobic residues present at the ends of the helices. The solved crystal structure of GlpT, in an inward facing conformation, shows the cytoplasmic ends of the helices facing away from each other. The diagram of structure comparison between GlpT's helices 1 and 7 can be seen in figure 4.1. A noticeable kink is seen towards the end of both helices. It is thought that this feature mediates conformational change during the rocker-type transport mechanism (122).

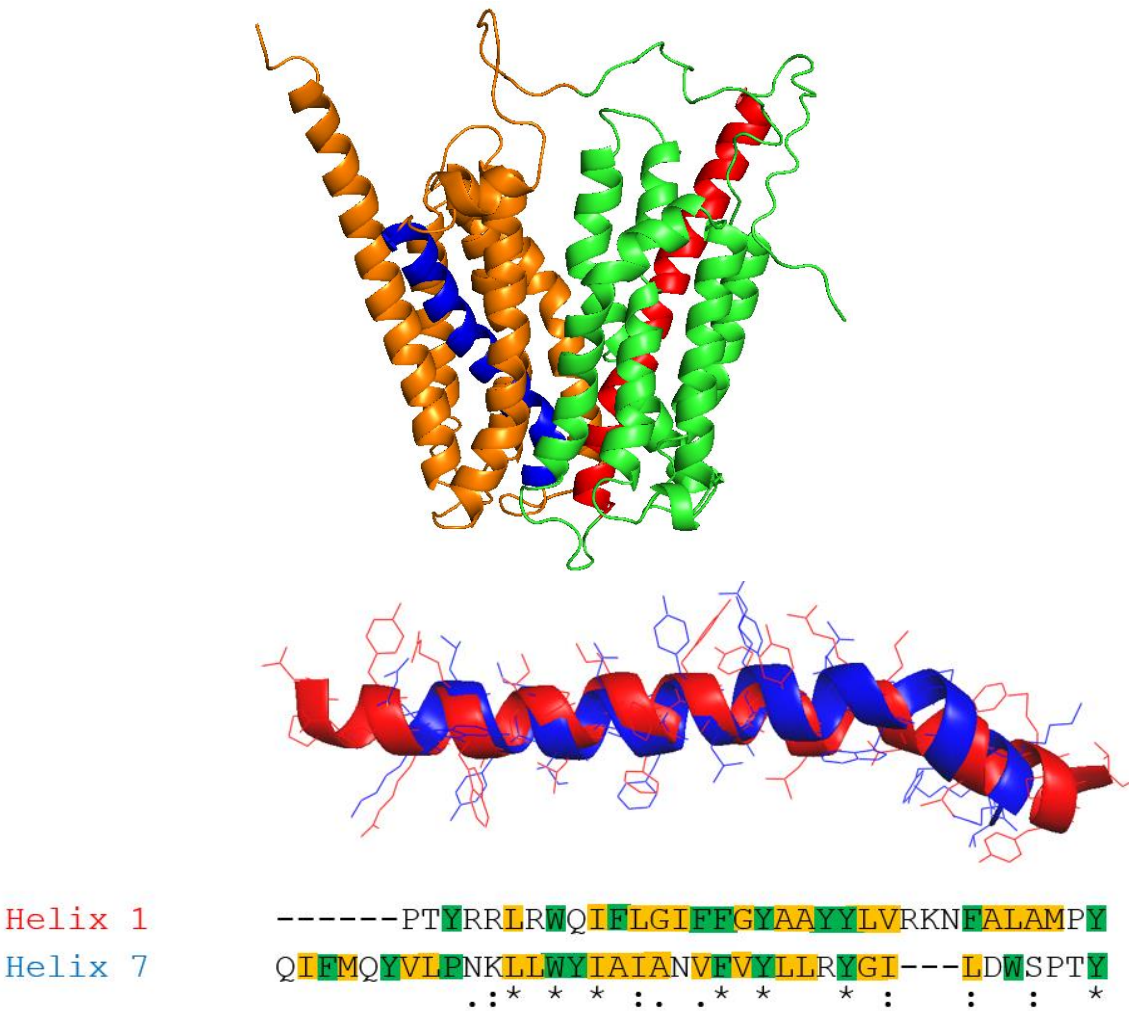


Figure 4.1 The structure and sequences of helices 1 and 7 aligned to each other

Top figure, the two domains of GlpT with helices 1 and 7 are highlighted and indicated by colour as follows: the N domain of the GlpT is coloured green, while the C domain is shown in orange, helix 1, shown in red, while helix 7 is displayed in blue. In the middle, the helices 1 and 7, represented as a ribbon structure, were superimposed against each other, using PyMOL software, to show similarities in their fold and length. In the bottom diagram, the peptide sequence alignment of helices 1 and 7 show differences in amino acid composition of the two helices. An asterisk indicates positions which have a single, fully conserved residue between two helices, a colon indicates conservation between amino acid groups with strongly similar properties, and finally a period sign indicates conservation between groups with weakly similar properties. The small hydrophobic residues are highlighted in yellow, while aromatic residues are highlighted in green. The secondary structure alignment was achieved by superimposing positions of α carbons, using data from protein database (1PW4, 97, Uniprot: P08194). Structures were rendered in PyMOL software. The peptide sequence alignment was done using the Clustal OMEGA sequence web service (https://www.ebi.ac.uk/Tools/msa/clustalo/).

4.1.3 Effect of mutations

Alanine is a small hydrophobic residue with a single methyl group attached. Substitution to alanine removes carbon chains past the β-carbon. Any functional group that may have contributed to the stability or function is thus eliminated. Alanine substitution methodology helps to identify contributions of amino acids containing large side chains (179). Substitution into glycine would also serve a similar purpose. However, glycine possesses only a single hydrogen as its side chain. Substitutions to alanine retain some of the original size as well as mimic the secondary structure with the benefit of not having any potential functional group. In membrane proteins, it is important to also consider the energetic contributions of the bilayer or the solubilising agent, when discussing the stabilising influence of the amino acid residue. Alternatively, another small amino acid can be introduced instead of alanine in order to conserve the size of the mutated residue. In general, apolar to apolar substitutions are more conservative than polar to apolar or aromatic to apolar change (138).

Multiple studies have shown that the folding pathway tends to remain largely unaffected, regardless of the positions of alanine substitutions (180). However, the changes in the free energy of unfolding, caused by an alanine mutation, cannot by directly associated with the equivalent energy that it takes to disrupt the protein structure. Both the position of the mutation and the surrounding environment contribute to the net energy change that results from the mutation. In order to make the mutations more informative, the number of potential factors that may influence the final energy change needs to be kept to a minimum. Thus, substitutions of large hydrophobic residues for similar smaller hydrophobic residues are much easier to analyse, as the main difference after the substitution is a reduction of the surface area (180).

4.1.4 Alanine mutations in GlpT

The alanine substitution sites were chosen based on the similarities of their position on both helices 1 and 7, and the surrounding environment they were found in. Only the mutations of hydrophobic residues were made, replacing larger hydrophobic amino acids with smaller hydrophobic side chains. The Y38A-Y266A mutation pair was based on its central location in the interface between two domains and its functional importance in mediating interactions with the substrates of GlpT. Reduction of the side chain size during the alanine substitution leads to creation of cavities between the helices, thus disrupting potential interactions that could potentially play a role during the protein folding event. These mutations are meant to test the targeted amino acids contributions to stability, as well as, protein's resistance to these disruptions.

The representation of locations of various alanine substitutions can be seen in figure 4.2. The structures of individual mutations and their respective environments are located in the supplementary appendix D. The residues Y38 and Y266 are located in the middle of both helices and are facing the central cavity, where they may interact with the GlpT substrates. Residues L26, I34, L254 and I258 are located on the cytoplasmic interface of the plasma membrane. These residues are facing towards peripheral helices 3 and 6, for L26 and I34, and helices 9 and 12, for L254 and I258. Both L44-L268 and V43-L267 are located in the middle of the helices 1 and 7. They are facing towards the interhelical cavity made by the central helices 4 and 10 as well as peripheral helices 3, 6, 9 and 12. The final set of mutated residues, L50-L273 and L54-L279, are positioned near the periplasmic interface. Residues L50 and L273 are facing the same direction towards helices 5 and 8 respectively. Finally, L54 and L279 are set opposite to each other towards helices 4 and 10, respectively. The positions of residues and the final protein structure *in vivo* may differ from the crystal structure.

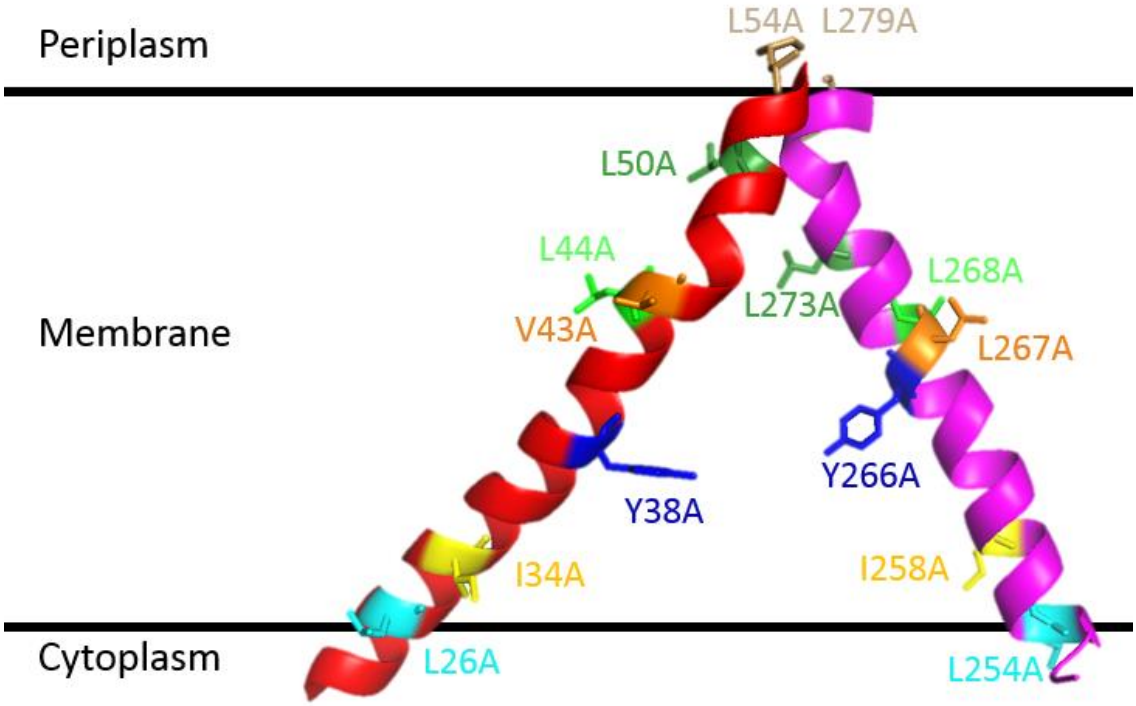


Figure 4.2 Diagram of alanine substitutions in the helices 1 and 7

The individual alanine mutation sites are shown above and are colour coded along with their respective mutation on the opposite helix. The mutations are listed from top to bottom in the following order: L54A-L279A, L50A-L273A, L44A-L267A, V43A-L268A, Y38A-Y266A, I34A-I258A, L26A-L254A.The more detailed diagrams of mutations and their locations within the protein can be found in appendix D. Structure was rendered in PyMOL software.

4.1.5 Choice of CD measurements over fluorescence

The method of measuring changes in structure upon denaturant unfolding was restricted to the use of far-UV CD spectroscopy. The previous measurements made by fluorescence, detailed in chapter 3, showed a complex aromatic environment, thus complicating interpretations of the results. Focus on the single global method in the unfolding assay, like the CD spectroscopy, is advantageous in detailed reporting of differences in stability and is therefore used in the following investigation.

4.1.6 Aims of the chapter

GlpT was found to undergo reversible denaturation. In this chapter the goal is to determine the free energy change of unfolding in the absence of denaturant in order to determine the differences in stability between helices 1 and 7 of GlpT and compare them to the stability of the WT. The free energy of unfolding was determined using established protocols of urea denaturation and refolding assay, previously introduced in chapter 3. The progress of unfolding with urea and temperature denaturation was monitored with the use of CD spectroscopy. Furthermore, every mutant was evaluated by control experiments, testing whether the alanine substitutions have hindered the ability of the protein to fully undergo the process of reversible denaturation. This was achieved by measurements of structure recovery upon dilution of denaturant and return of the original binding affinity towards substrate in the functional assay. In the end, the significant differences in stability of helices 1 and 7 will be discussed based on the alanine substitution data.

4.2 Results

4.2.1 Stability of the alanine mutants in urea

The protein stability and resistance to mutation are examined in GlpT's alanine mutants, which have been constructed with successive alanine substitutions of hydrophobic residues in helices 1 and 7. The progress of each alanine mutant unfolding in urea was followed by far-UV CD spectroscopy The intensity of the CD signal at 222 nm band was normalised to MRE and plotted against the urea concentration. A more accurate estimation of the unfolding would be gained by measuring both the amplitude of a much stronger absorption band at 190 nm and the negative bands at 210 and 222 nm. Together these measurements provide a detailed description of denatured protein's structure content. However, because of the high concentrations of urea needed to unfold alpha helical protein, the measurement at 190 nm is inaccessible due to the denaturant absorption within that area of the spectrum.

The CD spectra of folded and unfolded states, as well as unfolding curves from the urea denaturation assays, are shown for each alanine pair: L26A-L254A, I34A-I258A, Y38A-Y266A, V43A-L267A, L44A-L268A, L50A-L273A, L54A-L279A in the following figures 4.3, 4.4, 4.5, 4.6, 4.7, 4.8 and 4.9, respectively; all CD spectra presented were recorded at 1 nm intervals with 1 second acquisition time from 270 to 190 nm in the buffer solution and between 270 and 205 nm in urea. All scans were performed in 0.2 mm pathlength cell with 0.2-0.4 $mg.ml^{-1}$ protein concentration. The raw spectra were first processed in CDTool then plotted and converted to MRE in GraFit software. The data was fit to the two state curve in GraFit, the following fit parameters are displayed in Tables 4.1 and 4.2.

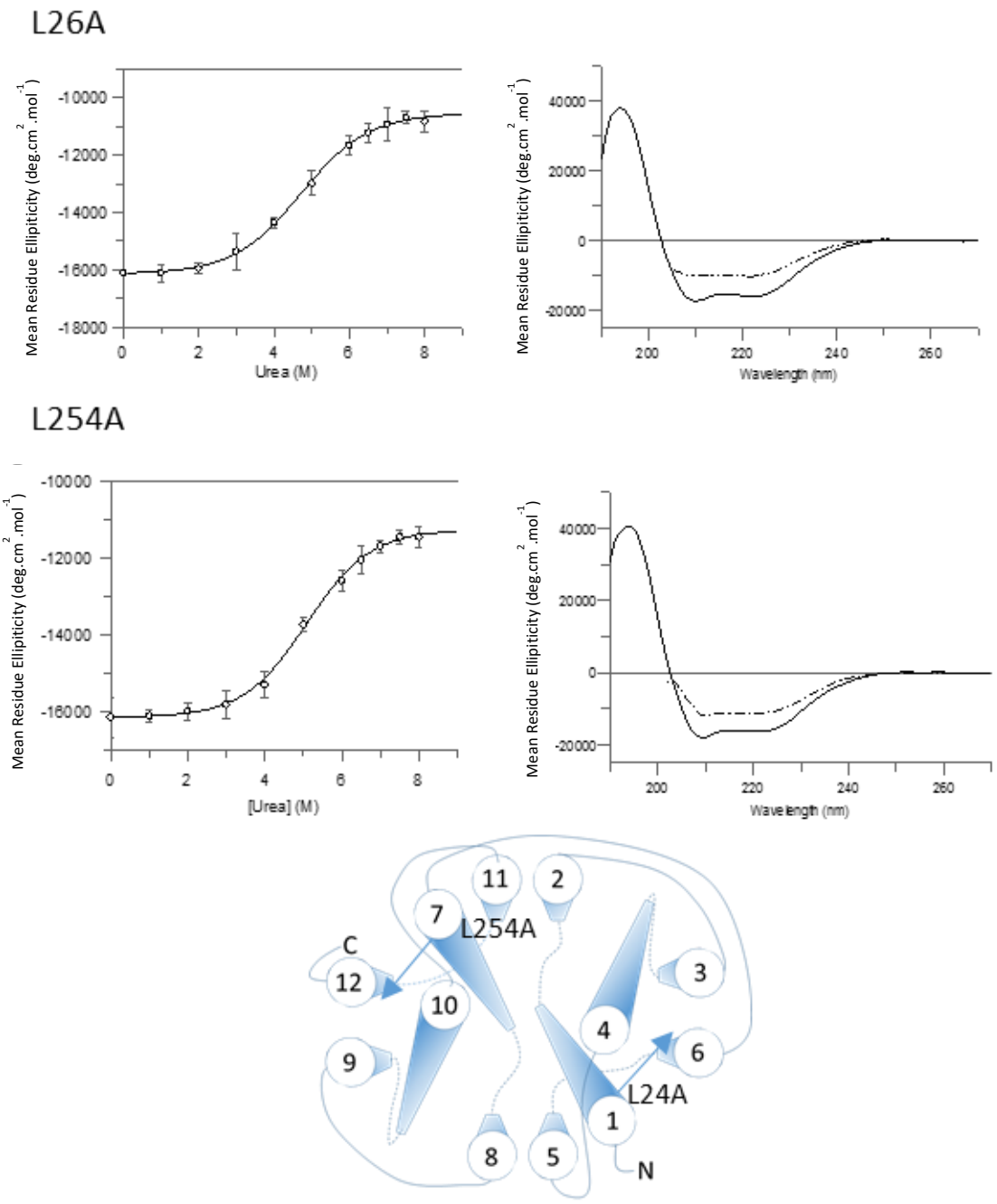


Figure 4.3 The L26A and L254A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of a L26A-L254A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for L26A and L254A mutant. Below, the position and orientation of mutated residues is shown as arrows.

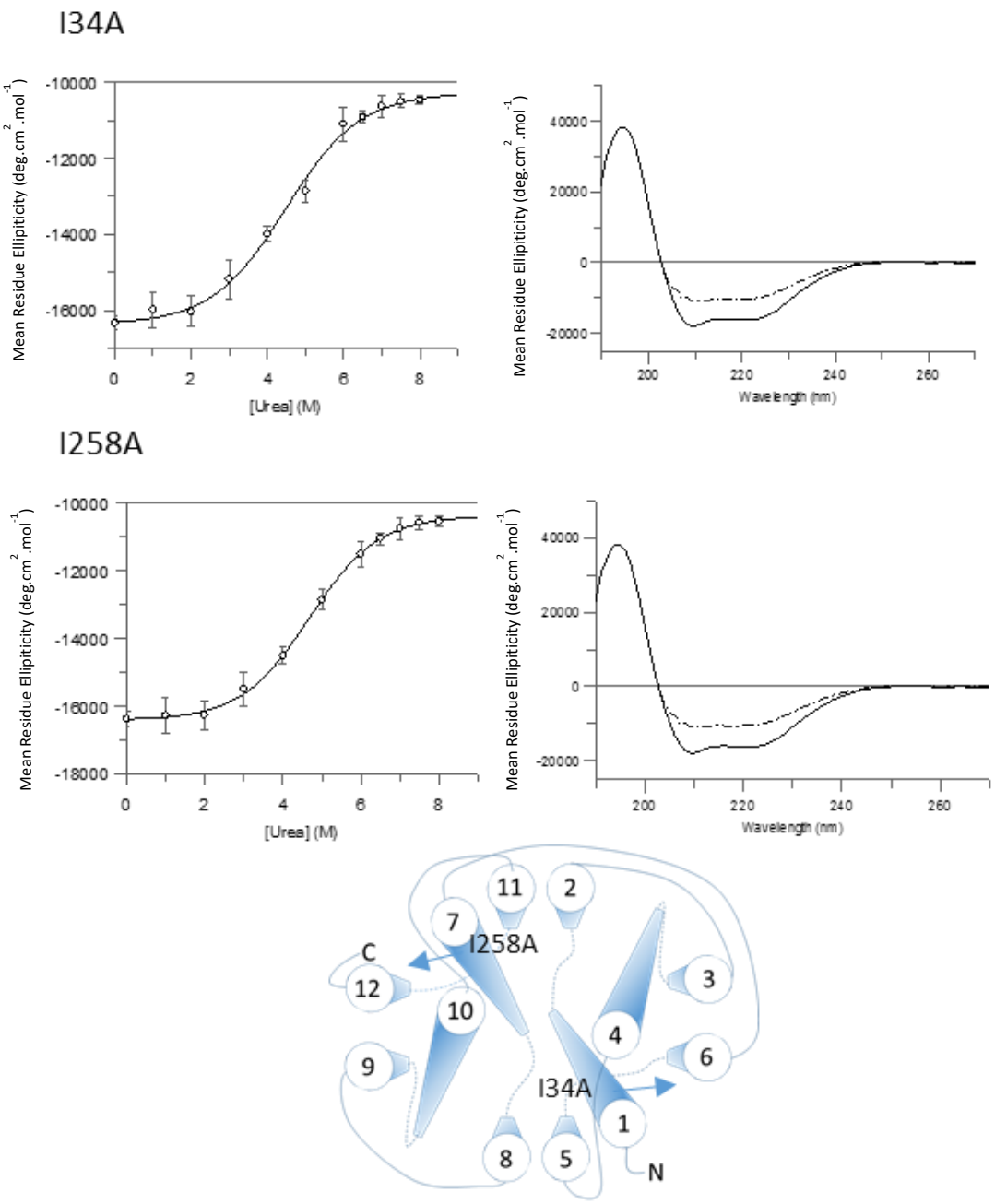


Figure 4.4 The I34A and I258A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of an I34A-I258A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for I34A and I258A mutant. Below, the position and orientation of mutated residues is shown as arrows.

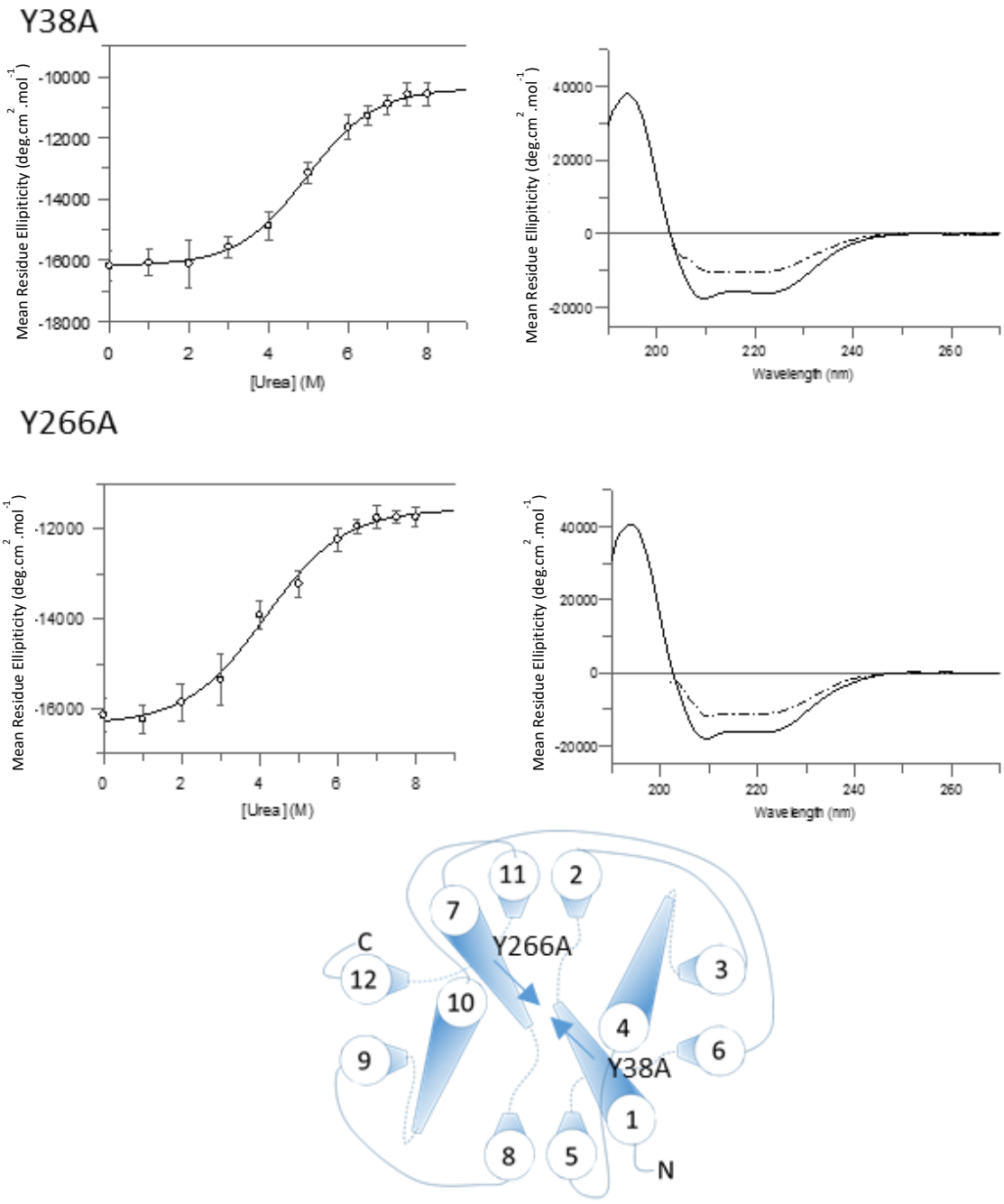


Figure 4.5 The Y38A and Y266A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of an Y38A-Y266A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for Y38A and Y266A mutant. Below, the position and orientation of mutated residues is shown as arrows.

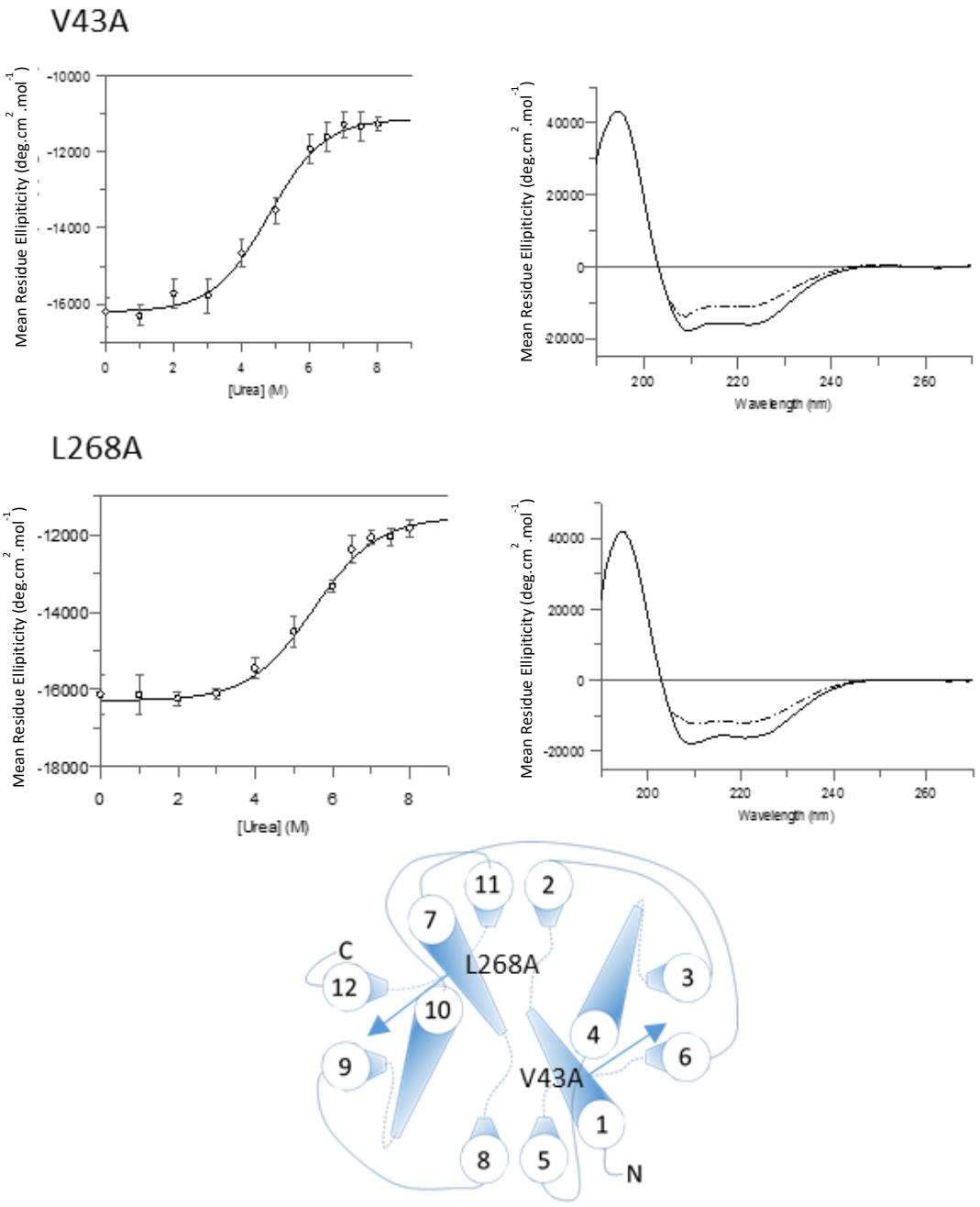


Figure 4.6 The V43A and L268A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of a V43A-L268A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for V43A and L268A mutant. Below, the position and orientation of mutated residues is shown as arrows.

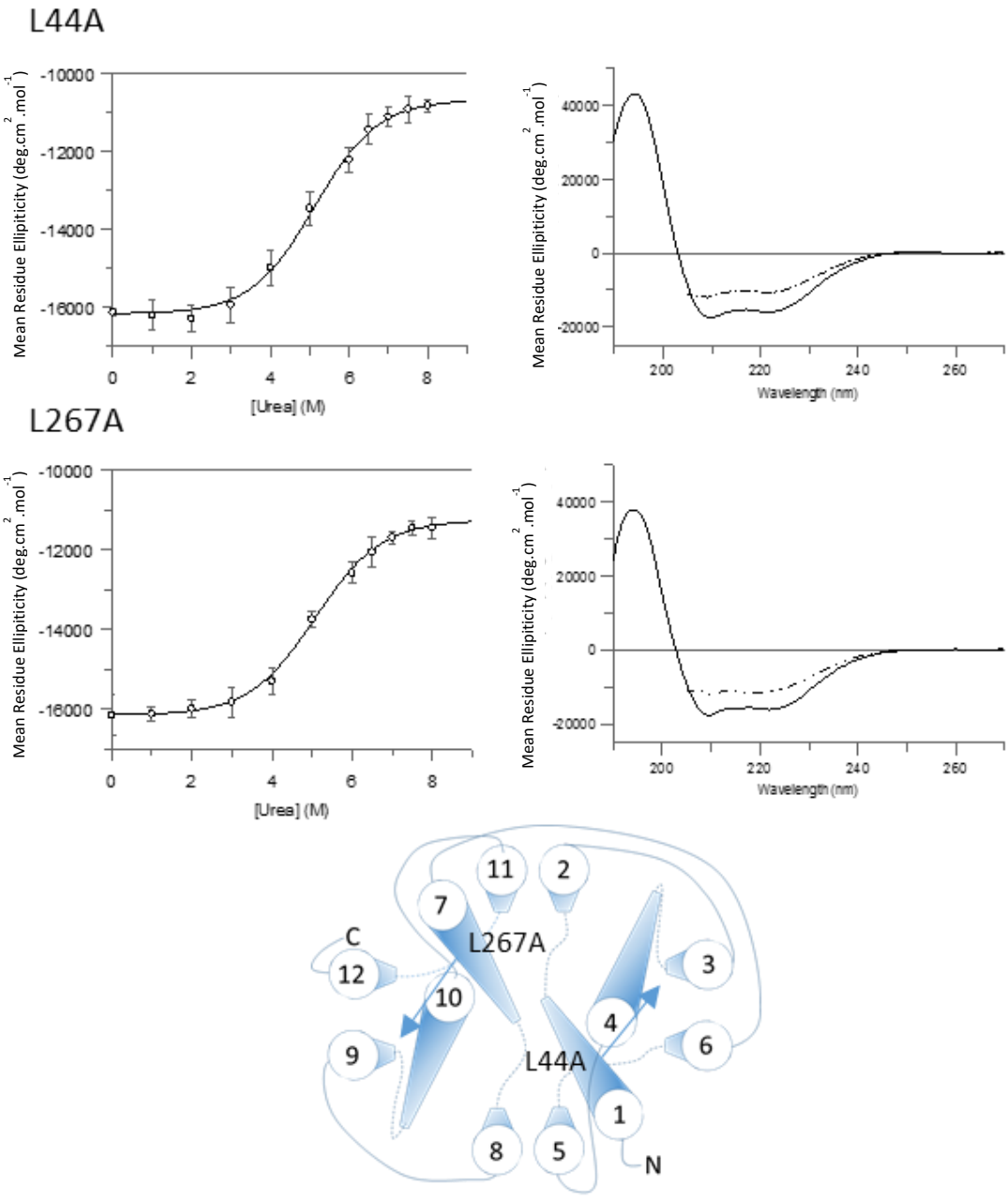


Figure 4.7 The L44a and L267A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of a L44A-L267A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for L44A and L267A mutant. Below, the position and orientation of mutated residues is shown as arrows.

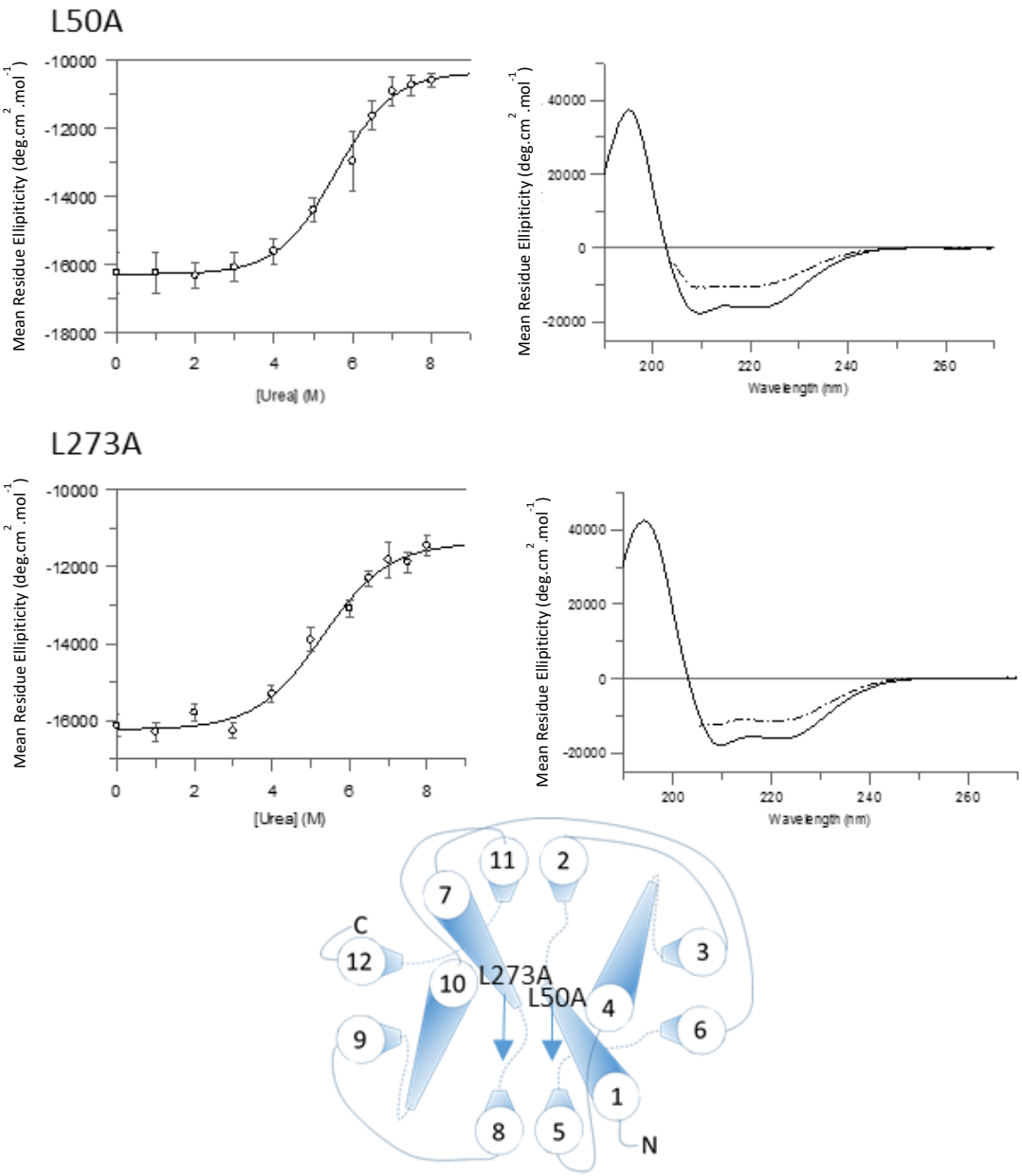


Figure 4.8 The L50A and L273A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of a L50A-L273A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for L50A and L273A mutant. Below, the position and orientation of mutated residues is shown as arrows.

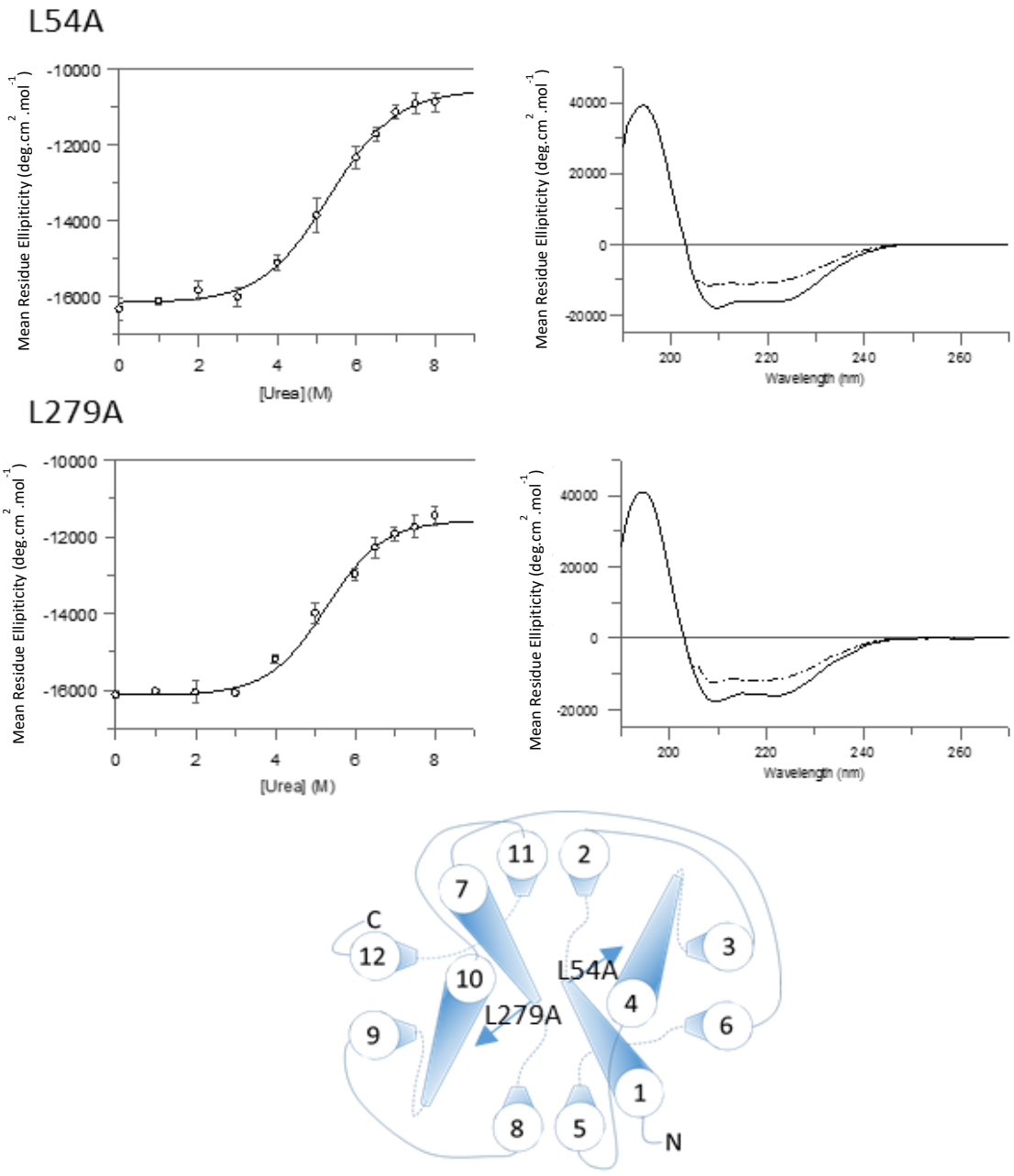


Figure 4.9 The L54A and L279A urea unfolding curves with their respective CD spectra

On the left, the decay of the 222 nm negative band is plotted against 0-8 M urea concentrations to show the unfolding steps of a L54A-L279A unfolding pair, when incubated with urea for 5 minutes. On the right, individual scans of the native state, (solid line), and denatured state in the 8 M urea, (dotted line), are shown for L54A and L279A mutant. Below, the position and orientation of mutated residues is shown as arrows.

Comparison of the CD spectra of alanine mutants shows no differences in the shape or magnitude of alpha helical bands in the far-UV spectrum (Figure 4.10). All minor differences are thought to be due to quality of protein preparation, rather than differences in the structural content.

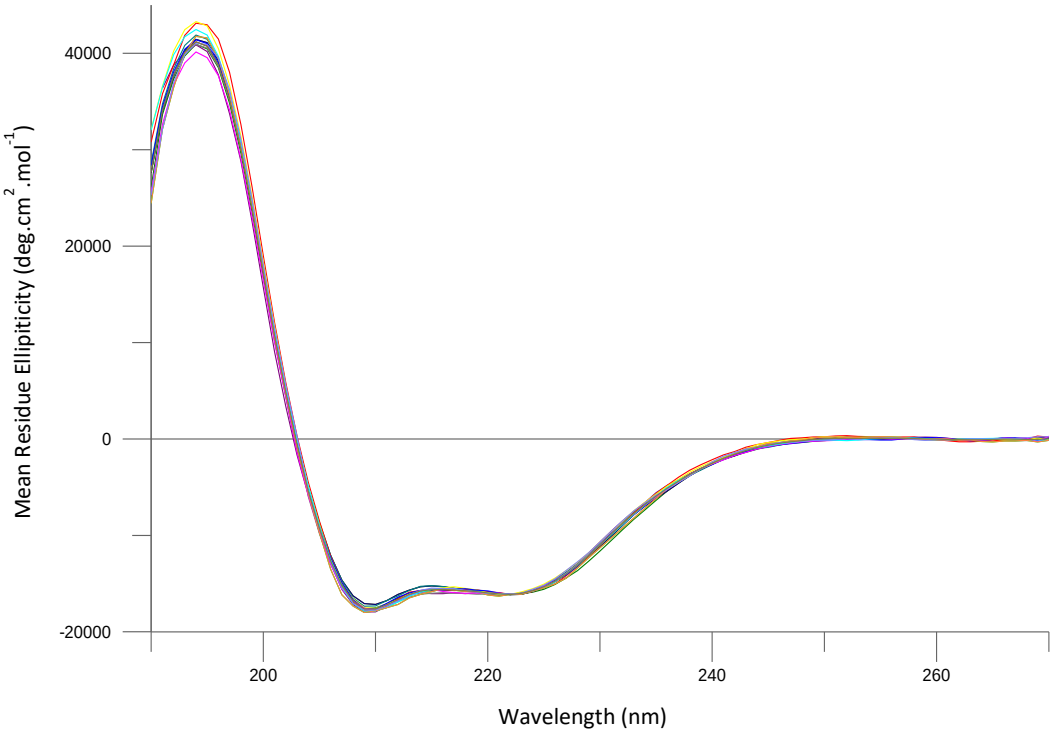


Figure 4.10 Overlay of all CD scans of folded alanine mutants and GlpT

CD scans of GlpT WT and alanine mutants were overlaid to show similar shape and magnitude of CD measurements. Spectra show only small differences in shape, which are due to differences of sample concentrations and noise of the measurement. Spectra were recorded in a 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 270 nm to 190 nm in buffer. The raw spectra from three repeated measurements was first averaged with the buffer scan subtracted, the baseline between 270 and 190 nm was zeroed in CDTool then plotted and converted to MRE in GraFit software.

All CD scans of folded protein were done with at least 5 repeats, and show similar shape and intensity within the range of -16150 deg.cm^2.dmol for 222 nm band, -18000 deg.cm^2.dmol^{-1} for 210 nm band and 40000 deg.cm^2.dmol^{-1} for 190 nm band, close to the CD measurements of the folded GlpT WT, showing no effect of the alanine substitutions on the GlpT's folded structure in DDM micelles. The comparison between

alanine mutant unfolding curves and the WT profile of unfolding is further discussed below in terms of the m-value and midpoint Cm (figure 4.11).

The effect of the alanine mutations on the unfolded state is shown as a MRE difference recorded at 222 nm, between the folded state in 0 M urea and unfolded state at 8 M. This is a total value of MRE reduction in helix as a consequence of denaturation and is displayed in table 4.1 Out of helix 1 mutations, I34A (figure 4.4), I258A (figure 4.4), Y38A (figure 4.5) and L50A (figure 4.8) mutants show an increase in the total MRE reduction compared to the WT and their equivalent alanine mutation on helix 7, with an average loss of 280 deg.cm^2.dmol^{-1} (6 % increase). Whereas, out of helix 7 mutations L254A (figure 4.3), V43A (figure 4.6), L268A (figure 4.6), L267A (figure 4.7), L273A (figure 4.8) and L279A (figure 4.9) all show an increased stability, with an average MRE change of 4820 deg.cm^2.dmol^{-1} (11% decrease in MRE lost) when compared to the WT.

In summary, out of all alanine mutations on helix 1, only V43A shows an increased stability of 5090 deg.cm^2.dmol^{-1} (7 % decrease in MRE lost), when compared to WT. Out of alanine mutations on helix 7, with the exception of I258A, all show a decrease in total MRE reduction due to unfolding. Overall, the majority of the mutations on helix 7 show a decrease in total MRE reduction due to the unfolding compared to the mutations located on helix 1, which led to a bigger loss of MRE when unfolded. Moreover, 6 helix 7 mutants have shown to have a significant effect on total MRE change upon unfolding, compared to 4 helix 1 mutants. This shows helix 7 has a higher sensitivity to mutations than helix 1.

Construct	MRE change at 222 nm ($deg.cm^2.dmol^{-1}$)	Ala_{MRE} - WT_{MRE} at 222 nm ($deg.cm^2.dmol^{-1}$)	Increase or decrease of total MRE lost due to denaturation compared with the WT
WT	5460 ± 200		
L26A	5630 ± 150	170	No change
L254A	4910 ± 150	-550	Decrease
I34A	6000 ± 200	540	Increase
I258A	5830 ± 80	370	Increase
Y38A	5860 ± 130	400	Increase
Y266A	5010 ± 300	-450	No change
V43A	5090 ± 150	-370	Decrease
L268A	4710 ± 240	-750	Decrease
L44A	5490 ± 140	30	No change
L267A	4720 ± 150	-740	Decrease
L50A	5940 ± 170	480	Increase
L273A	4850 ± 260	-610	Decrease
L54A	5600 ± 230	140	No change
L279A	4730 ± 210	-730	Decrease

Table 4.1 Comparison of structure changes alanine substituted GlpT in 8 M urea

The loss of structure was monitored by a decay of 222 nm negative band. The MRE change at 222 nm was calculated by subtracting an MRE value recorded at 0 M concentration of urea with a value recorded at 8 M urea, giving the total amount of MRE reduction in helix content as a result of unfolding. The difference between Ala mutants and WT was calculated by subtracting the total MRE value of WT against each of the mutants, so that negative values mean loss of stability and positive values indicate a gain of stability. The mean MRE values used to calculate the differences are the result of at least 5 experiments and come from at least 2 different protein purifications.

The measurements of the mutants located close to the periplasmic side of the membrane show a consistent profile of the protein unfolding, similarly to WT. For the majority of the structures, denaturation is observed after 3 M point, with the midpoint of the denaturation curve reached at 4.5 M urea concentration. Therefore, the resistance to the urea unfolding, indicated by the midpoint value, remains within the same magnitude, compared to WT. The calculated midpoint concentrations are shown in figure 4.11, along with the gradients of the transition phase. There seems to be a certain trend between an increasing mid-point value and the proximity of the alanine mutation site towards the side of the protein facing the periplasmic interface of the membrane in *E.coli*. The second parameter of the ΔGu^{H2O} calculation, the gradient of the transition phase m_u varies between different alanine mutants and shows larger errors, contributing to the overall error of the ΔGu^{H2O}.

Based on the mean $\Delta\Delta Gu^{H2O}$ comparison against the ΔGu^{H2O} WT ($\Delta\Delta Gu^{H2O}WT = \Delta Gu^{H2O}$ WT - ΔGu^{H2O} Ala) the most destabilising mutations, with change smaller than -0.2 kcal/mol were located on helix 1 (V43A, L44A, L50A). There was only one mutation on helix 7, that was found to be strongly destabilising (I258A). The mutations with $\Delta\Delta Gu^{H2O}$ above 0.2 kcal/mol were found to be L268A and L273A, both located on helix 7 and L54A, located on helix 1. The remaining mutations minimally altered the stability and are listed as: L26A, L254A, I34A, Y38A, Y266A, L267A and L279A. However, with the larger errors associated with the results it is difficult to fit the data accurately and make subsequent analysis. Although helix 1 shows a higher frequency of destabilising mutations, than helix 7, whereas helix 7 has more stabilising mutations, it is difficult to mark these results as significant due to a high error associated with the ΔGu^{H2O} calculations.

If the folding of the protein can be described by a two state reaction and the relationship between the unfolding and refolding is linearly dependent on the concentration of the

denaturant used, therefore it is possible to determine ΔGu^{H2O} by extrapolation of the data to the y-axis. The linear fits of transformed data are shown in figure 4.12, presenting the linear dependence of ΔGu on the denaturant concentration. The following results and parameters from the linear fit are summarised in table 4.3. When the reversible thermodynamic equilibrium is established, either sigmoidal two state plots or linear plots can be used in calculating the ΔGu^{H2O} value.

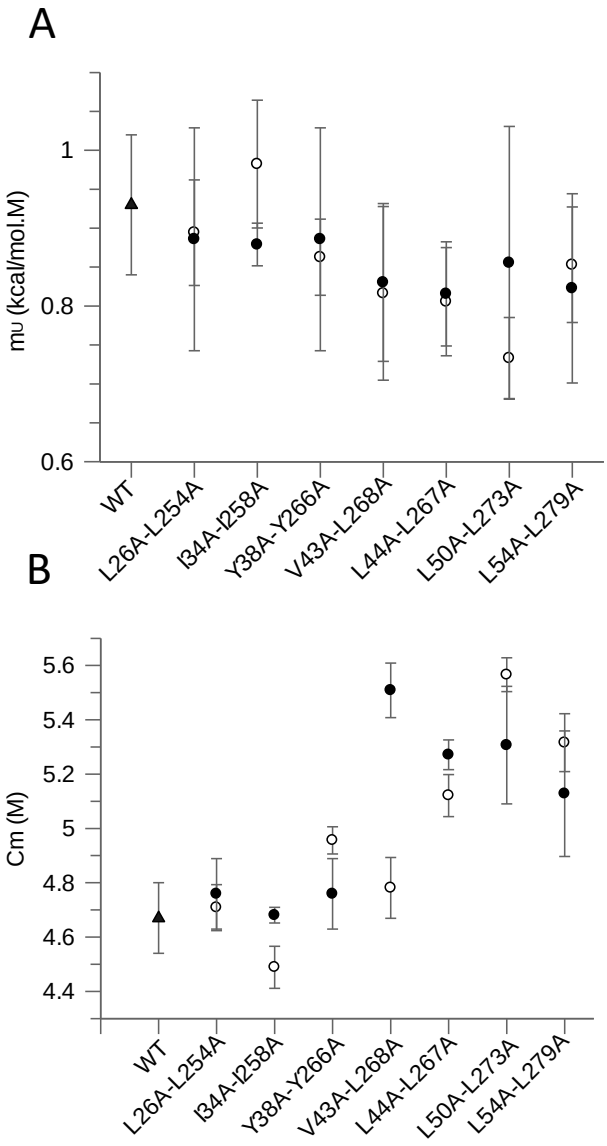


Figure 4.11 Differences in Cm and m-value between mutations on helix 1 and 7

The gradient, m_u and midpoint value, Cm calculated from the sigmoidal plots fitted to alanine mutants unfolding data are shown in panels A and B, respectively, with the WT values for comparison. Mutations on helices 1 and 7 are represented by white and black dots, respectively, with the WT shown as the black triangle. The data from the alanine urea unfolding experiments

were fitted to the two state curve in GraFit. The following fit parameters and resulting ΔGu^{H2O} are summarised in Table 4.2. Errors shown are from the standard error of the mean.

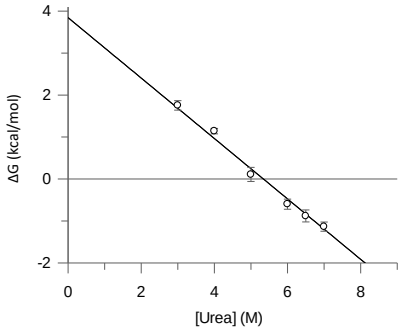
The error bars for the m_U^{H2O} and Cm were taken from the standard error of the fit and were used to calculate upper and lower bounds for the error of ΔGu^{H2O}. Similarly, the error of the m_U^{H2O} and ΔGu^{H2O} calculated from the linear transformation was taken from the fit. Although care was taken to reduce all errors in the measurements, unavoidable inaccuracy in protein concentration measurements as well as small dissimilarities in protein sample preparations have contributed to the error in CD measurements, which then lead to an error with the calculated ΔG values. To confirm whether the differences between ΔG values are real, further experiments are needed, as well as, extended controls of the protein reproducibility.

Construct	Cm (M)	ΔWT Cm	ΔAla Cm	m_U^{H20}	ΔGu^{H2O} kcal.mol^{-1}	$\Delta\Delta Gu^{H2O}$ WT	$\Delta\Delta Gu^{H2O}$ Ala
WT	4.67 ± 0.13			0.93 ± 0.09	4.3 ± 0.5		
L26A	4.71 ± 0.08	-0.04	-0.05	0.89 ± 0.07	4.2 ± 0.4	-0.1	0
L254A	4.76 ± 0.13	-0.09		0.89 ± 0.14	4.2 ± 0.8	-0.1	
I34A	4.49 ± 0.08	0.18	-0.19	0.98 ± 0.08	4.4 ± 0.4	0.1	0.3
I258A	4.68 ± 0.03	-0.01		0.88 ± 0.03	4.1 ± 0.2	-0.2	
Y38A	4.96 ± 0.05	-0.29	0.20	0.86 ± 0.05	4.3 ± 0.3	-0.1	0.1
Y266A	4.76 ± 0.13	-0.09		0.88 ± 0.14	4.2 ± 0.7	-0.1	
V43A	4.78 ± 0.10	-0.11	-0.73	0.81 ± 0.10	3.9 ± 0.5	-0.4	-0.7
L268A	5.51 ± 0.10	-0.84		0.83 ± 0.10	4.6 ± 0.5	0.3	
L44A	5.12 ± 0.08	-0.45	-0.15	0.80 ± 0.07	4.1 ± 0.4	-0.2	-0.2
L267A	5.27 ± 0.05	-0.60		0.81 ± 0.07	4.2 ± 0.4	-0.1	
L50A	5.58 ± 0.06	-0.91	0.27	0.73 ± 0.05	4.1 ± 0.3	-0.2	-0.5
L273A	5.31 ± 0.22	-0.64		0.85 ± 0.18	4.5 ± 0.9	0.2	
L54A	5.31 ± 0.11	-0.64	0.18	0.85 ± 0.07	4.5 ± 0.5	0.2	0.4
L279A	5.13 ± 0.23	-0.46		0.82 ± 0.12	4.2 ± 0.8	0.1	

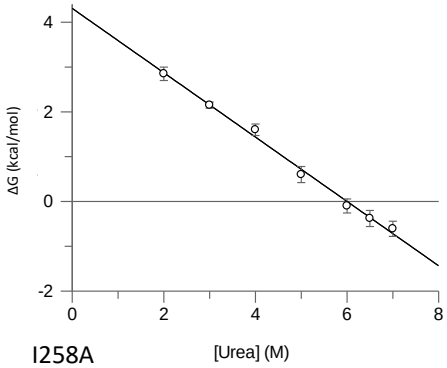
Table 4.2 Summary of ΔGu^{H2O} values for WT and Ala mutants from two state plots

The free energy values for alanine mutants and WT were calculated from sigmoidal fits to CD unfolding data, errors are taken from the standard error of the fit. The $\Delta\Delta Gu^{H2O}$ was calculated as a ΔGu^{H2O} WT - ΔGu^{H2O} Ala difference, while $\Delta\Delta Gu^{H2O}$ Ala is described as a helix 1 ΔGu^{H2O} – helix 7 ΔGu^{H2O}

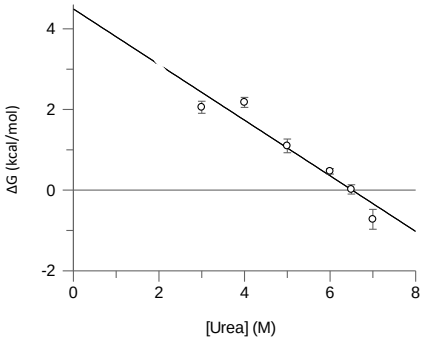
L26A



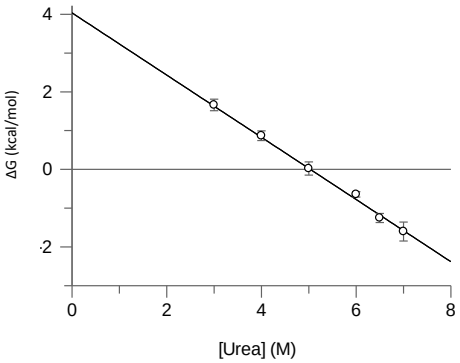
L254A



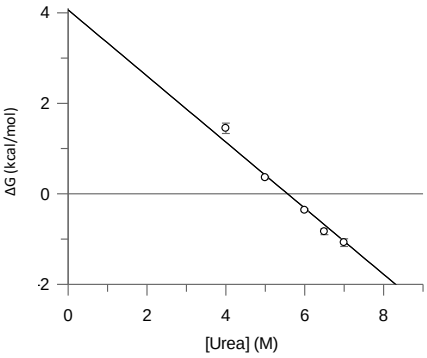
I34A



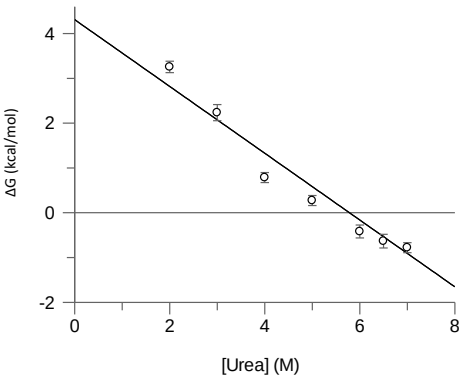
I258A



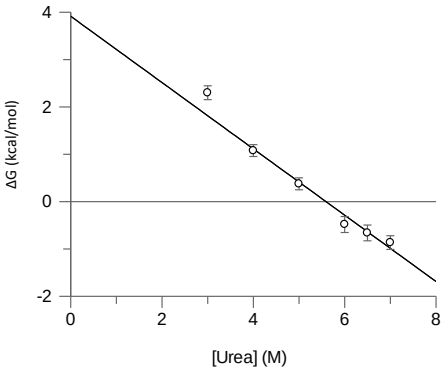
Y38A



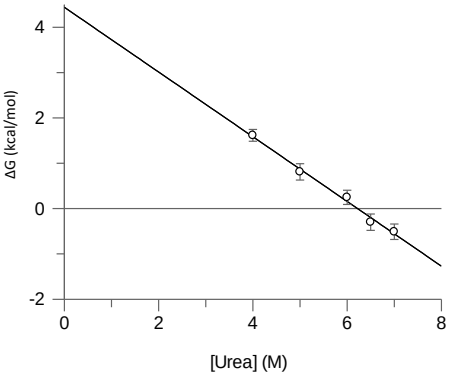
Y266A



V43A



L268A



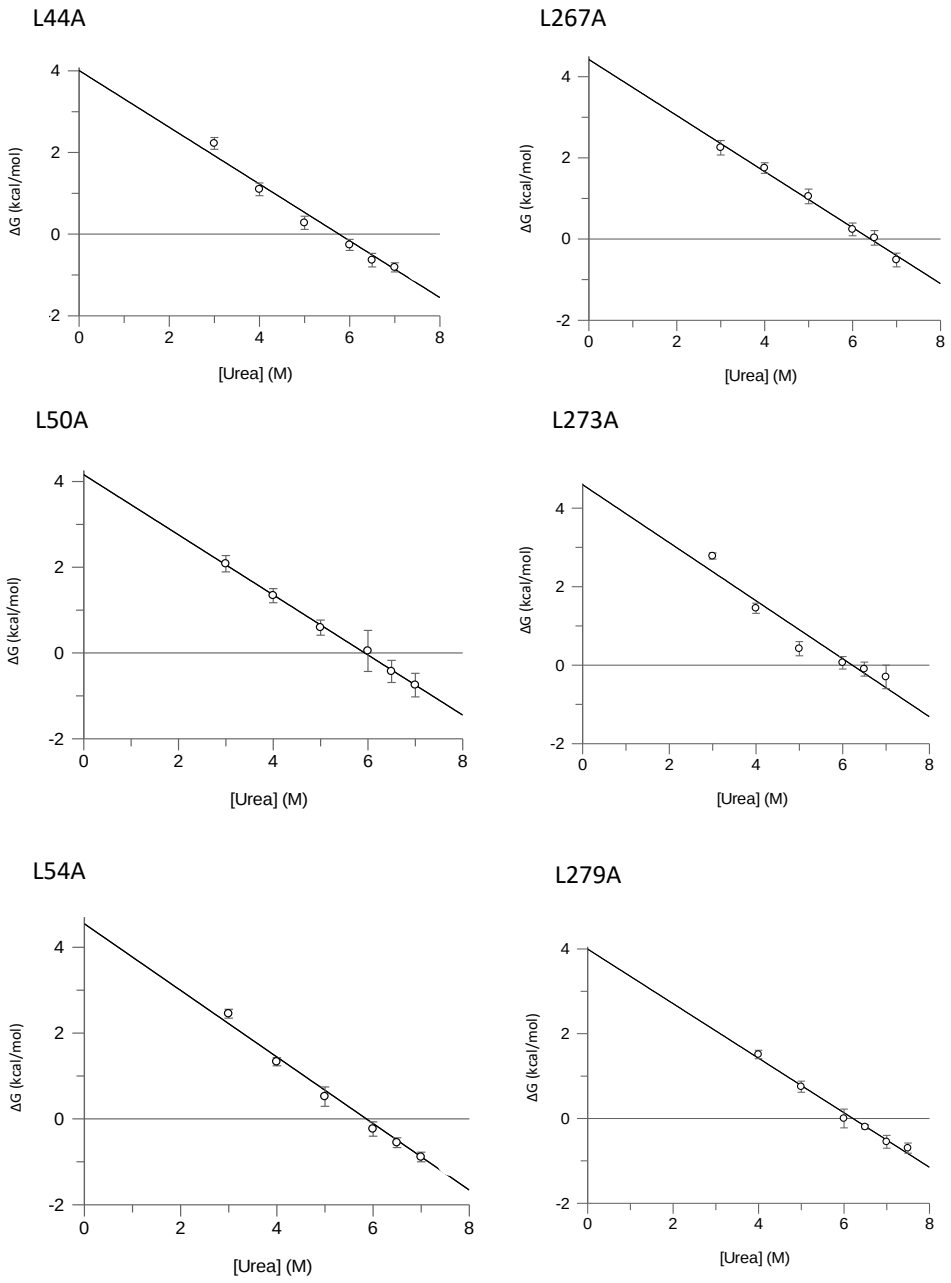


Figure 4.12 Linear fits of alanine mutant unfolding curves measured by CD spectroscopy

The progress of alanine mutations unfolding, measured by CD, is plotted against urea concentrations. The ΔGu^{H2O} value is calculated from ΔG extrapolation of the y-axis intercept. The fit parameters used in this analysis are summarised in Table 4.3. CD data was recorded in a 0.2 mm pathlength cell at 1 nm intervals with 1 second acquisition time. The raw spectra were processed in CDTool then converted to ΔG and plotted in GraFit software.

Construct	m_U^{H20}	ΔGu^{H2O} (kcal.mol^{-1})	$\Delta\Delta Gu^{H2O}$ WT	$\Delta\Delta Gu^{H2O}$ Ala
WT	-0.85 ± 0.07	3.8 ± 0.6		
L26A	-0.72 ± 0.06	3.9 ± 0.3	0.1	-0.4
L254A	-0.73 ± 0.05	4.3 ± 0.4	0.5	
I34A	-0.69 ± 0.07	4.5 ± 0.5	0.7	0.5
I258A	-0.80 ± 0.03	4.0 ± 0.2	0.2	
Y38A	-0.78 ± 0.09	4.1 ± 0.9	0.3	-0.2
Y266A	-0.75 ± 0.08	4.3 ± 0.5	0.5	
V43A	-0.70 ± 0.06	3.9 ± 0.3	0.1	-0.6
L268A	-0.68 ± 0.06	4.4 ± 0.4	0.6	
L44A	-0.70 ± 0.06	4.0 ± 0.3	0.2	--0.4
L267A	-0.66 ± 0.03	4.4 ± 0.3	0.6	
L50A	-0.70 ± 0.02	4.2 ± 0.2	0.4	-0.4
L273A	-0.74 ± 0.08	4.6 ± 0.6	0.8	
L54A	-0.78 ± 0.04	4.6 ± 0.3	0.8	0.6
L279A	-0.65 ± 0.04	4.0 ± 0.3	0.2	

Table 4.3 Summary of ΔGu^{H2O} values for WT and Ala mutants from linear plots

The free energy values for alanine mutants and WT were calculated from linear fits to CD unfolding data, shown in figure 4.11. Errors are taken from the standard error of the fit. The $\Delta\Delta Gu^{H2O}$ was calculated as ΔGu^{H2O} Ala - ΔGu^{H2O} WT, while $\Delta\Delta Gu^{H2O}$ Ala is described as helix 1 ΔGu^{H2O} – helix 7 ΔGu^{H2O}.

4.2.2 Reversibility of urea induced unfolding

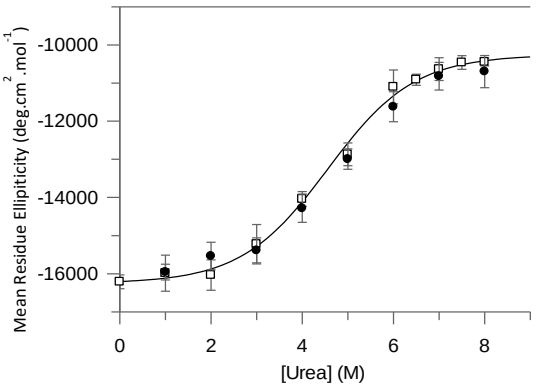
The conditions of the two state model, under which alanine ΔG values were calculated, state that the folded and unfolded protein exist in a reversible equilibrium. Therefore, to prove this, the unfolded state needs to be shown as recoverable. Similar to principles laid out in chapter 3, successful refolding of alanine mutants is assessed by regaining the original CD spectrum before the unfolding and the recovery of P5P ligand affinity.

First, the protein is unfolded by 5 minute incubation in urea, then the sample is refolded by rapid dilution, thus eliminating high concentration of the urea. Figure 4.13 shows the return of the CD signal at different urea concentrations, measured by MRE recorded at 222 nm and plotted at corresponding concentrations of urea. The refolding curves of all alanine mutants show complete recovery of protein structure upon refolding from the unfolded state. Both refolding and unfolding curves are superimposable, with slight deviation at few points, located at the 4 M and 5 M urea concentrations at the transition phase on the refolding curve. These deviations could possibly be caused by an error in accurate estimation of protein concentration and reduced CD signal due to a necessary but sizeable dilution of the protein sample.

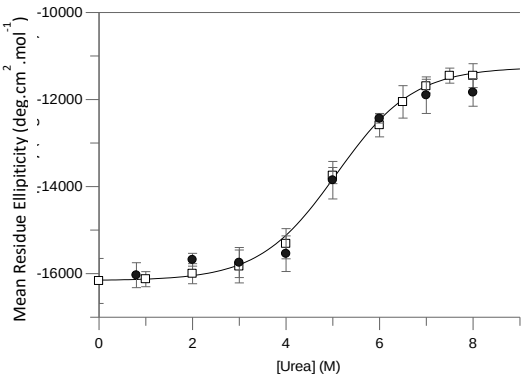
A ligand binding assay was introduced to further test the reversibility of WT and alanine mutants unfolding. The refolding assay measured by CD informs of the global return of the secondary structure, but does not provide detail on the tertiary contacts. The ligand binding assay measures the return of the ligand affinity to the refolded sample. Out of GlpT ligands tested in chapter 3, the P5P was found to be the best ligand to test the return of the binding affinity in the refolding assay, due to its large and measureable effect on the CD signal.

The binding was shown to follow closely with the single binding site model and can be measured as an increase in CD signal upon P5P binding. The saturation binding curves are shown in figure 4.14, with the summary of binding constants presented in table 4.5. The binding constants for both folded and refolded samples were all found to be close to each other, confirming the results shown by the refolding assay. The binding assay not only tells of recovery of binding affinities, but also the effect of the mutations on the binding of P5P. Out of all alanine mutations, only the V43A mutation led to a small change of binding affinity. Surprisingly, the mutations Y38A and Y266, located near the binding pocket, did not affect the binding constant.

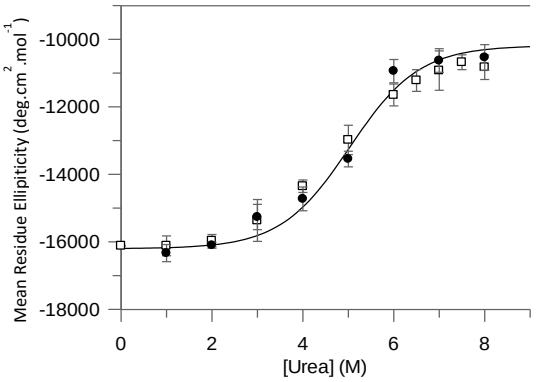
L26A



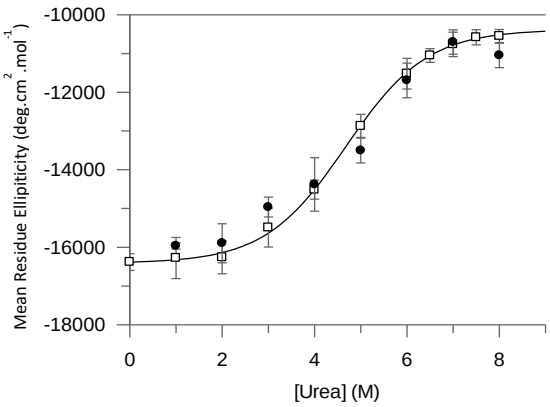
L254A



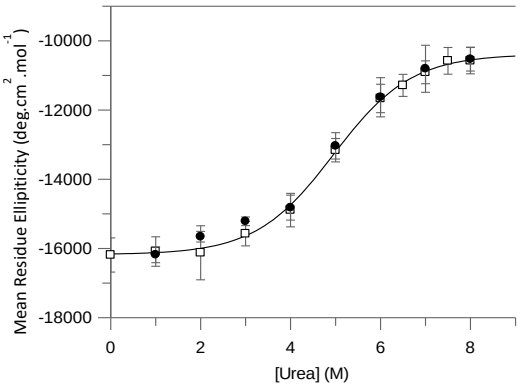
I34A



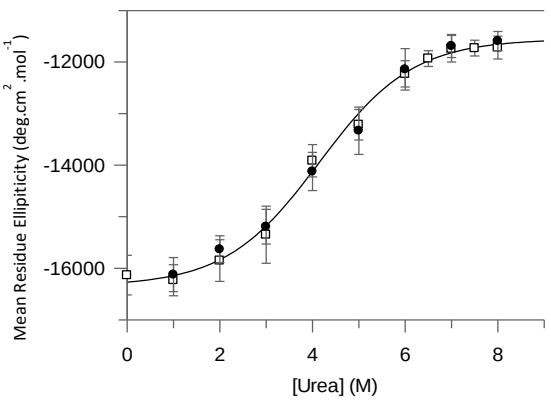
I258A



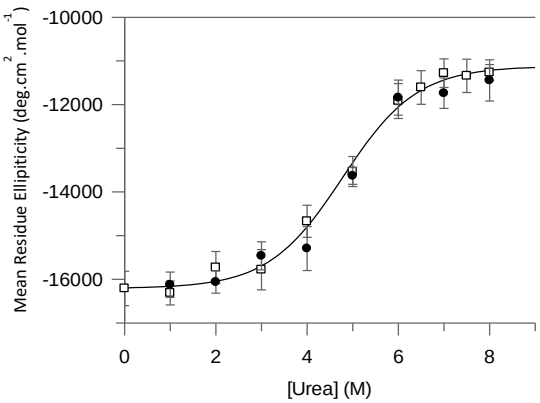
Y38A



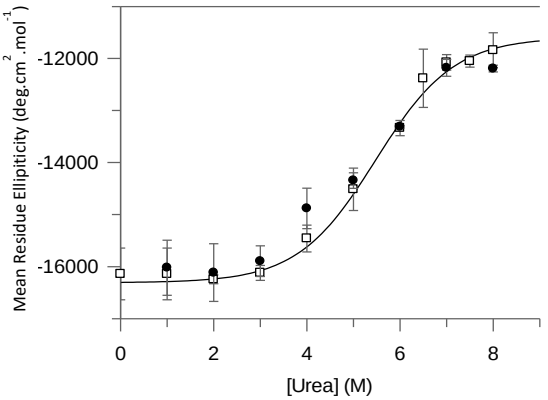
Y266A



V43A



L268A



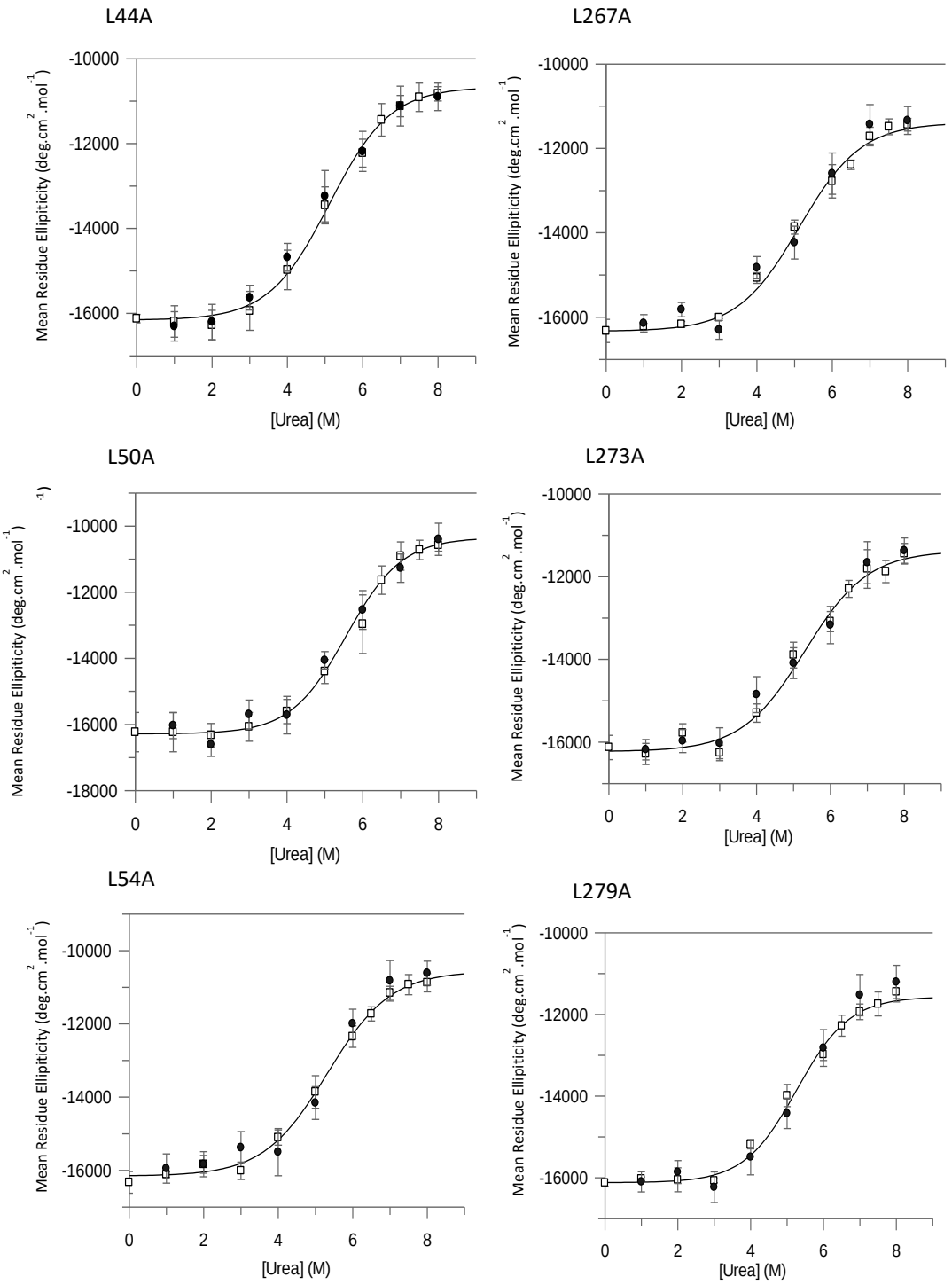


Figure 4.13 Refolding curves of alanine mutants

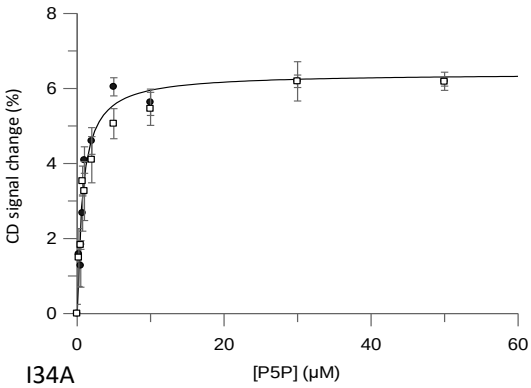
The extent of structure recovery was monitored at each urea concentration by CD spectroscopy. The unfolding data with its sigmoidal fits are represented by white squares and the refolding curves, shown by black dots. Both curves are superimposed to compare the similarities between unfolding and refolding data. The error bars are due to at least 3 repeats of each data point. The data was fit to two state curve in GraFit and the following fit parameters are displayed in Table 4.4.

Construct	m_f^{H20}	Cm	ΔGf^{H2O} (kcal/mol)
WT	0.85 ± 0.08	4.96 ± 0.05	4.2 ± 0.5
L26A	0.82 ± 0.15	5.03 ± 0.22	4.1 ± 0.9
L254A	0.72 ±0.08	5.10 ± 0.12	3.7 ± 0.5
I34A	0.94 ± 0.10	4.49 ± 0.07	4.2 ± 0.5
I258A	0.88 ± 0.09	4.68 ± 0.03	4.1 ± 0.4
Y38A	1.02 ± 0.16	5.02 ± 0.12	5.1 ± 0.9
Y266A	1.07 ± 0.16	4.82 ± 0.10	5.2 ± 0.9
V43A	0.96 ± 0.18	4.93 ± 0.12	4.7 ± 0.9
L268A	0.92 ± 0.14	5.35 ± 0.11	4.9 ± 0.8
L44A	1.03 ± 0.08	4.85 ± 0.08	5.0 ± 0.5
L267A	0.81 ± 0.34	5.23 ± 0.04	4.2 ± 0.5
L50A	0.96 ± 0.30	5.57 ± 0.03	5.3 ± 0.5
L273A	0.92 ± 0.30	5.49 ± 0.03	5.1 ± 0.5
L54A	0.85 ± 0.07	5.31 ± 0.11	4.5 ± 0.5
L279A	0.84 ± 0.12	5.27 ± 0.23	4.4 ± 0.8

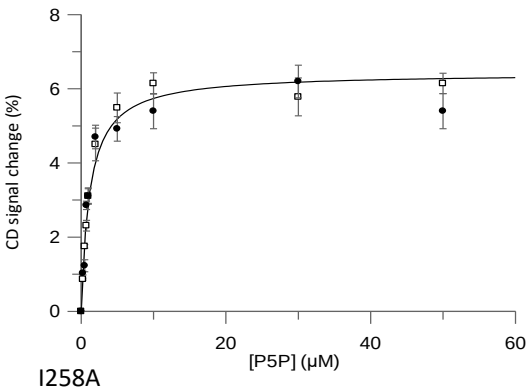
Table 4.4 Summary of the ΔGf^{H2O}, m_f and Cm values for the refolded alanine mutants and WT

Parameters for alanine constructs were calculated by fitting a sigmoidal equation to each refolding data set from the urea-induced unfolded state, shown in figure 4.13. Errors taken from the standard error of the fit.

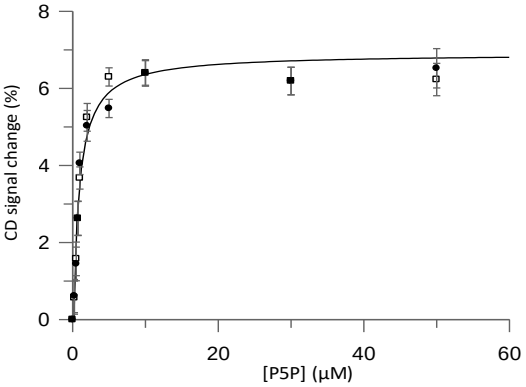
L26A



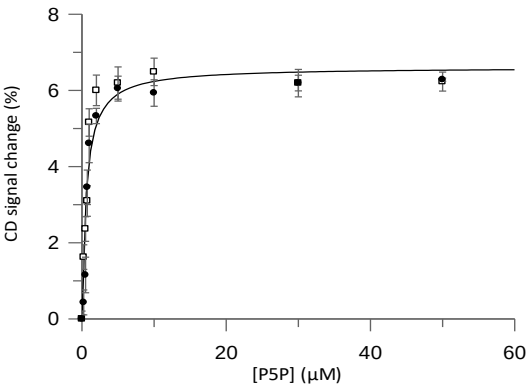
L254A



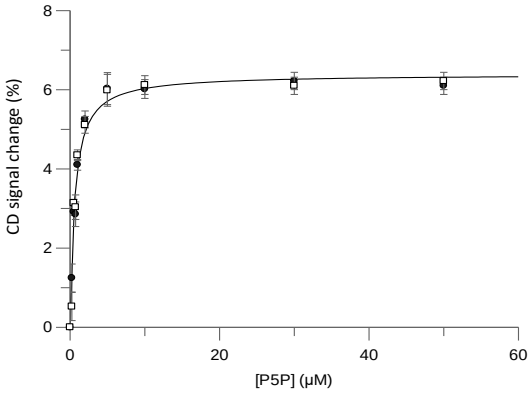
I34A



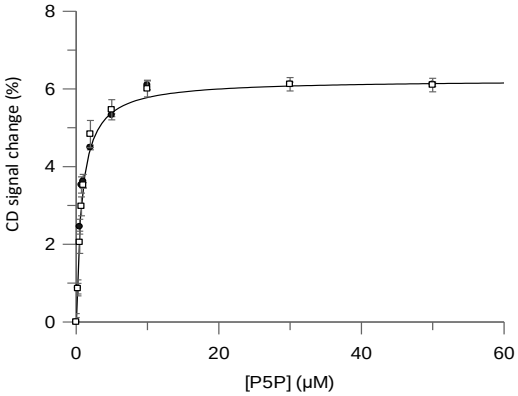
I258A



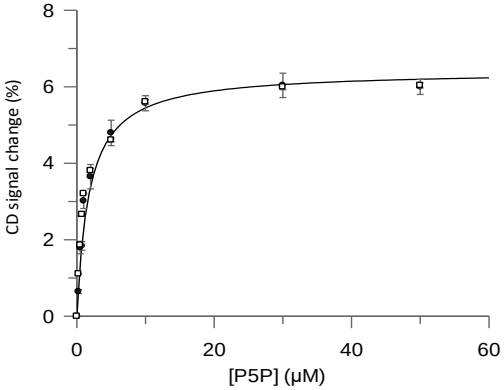
Y38A



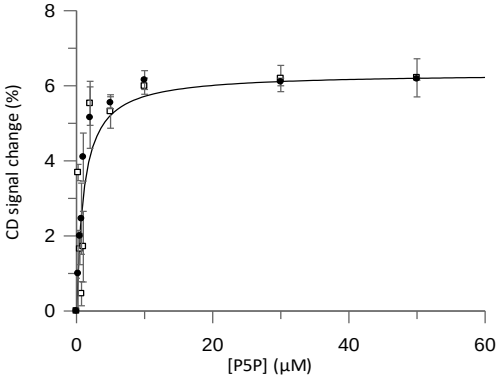
Y266A



V43A



L268A



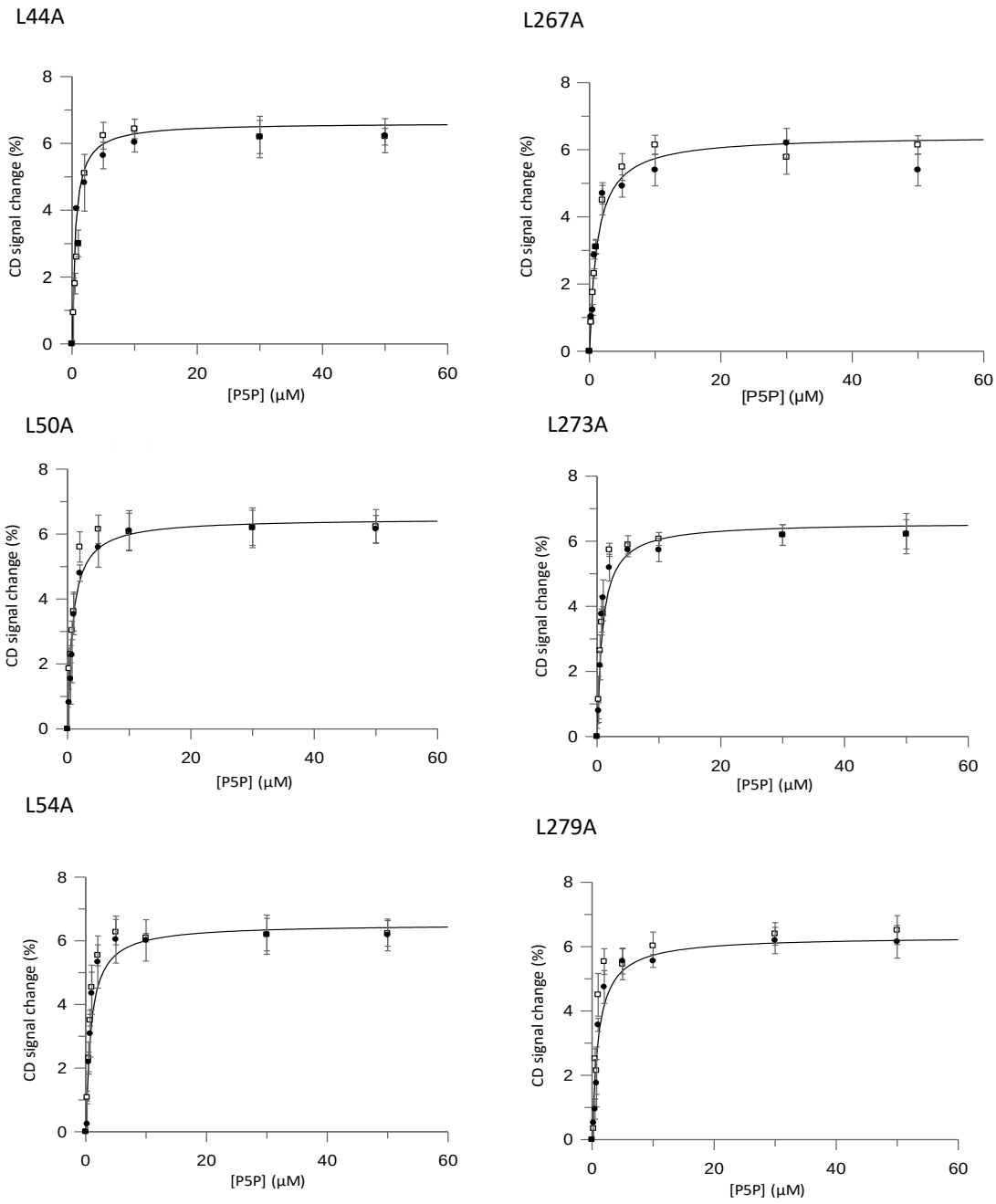


Figure 4.14 Saturation binding curves for folded and refolded GlpT alanine mutants

P5P-induced changes, measured by an increase in CD signal are plotted as a percentage of structure changed against corresponding P5P concentration. The binding curve of a folded sample, with its single binding fit, is represented by black dots, whereas refolded protein binding is marked as white squares. The fitting of refolded data was omitted for clarity. The binding plots show a similar extent of binding between folded and refolded samples. All scans were performed in a 2 mm pathlength cell in an in-house Aviv Circular Dichroism Spectrophotometer, Model 410. The raw spectra were first processed in CDTool then plotted and converted to a percentage change in GraFit software. The data was fit to the single binding site model in GraFit with the binding constant values displayed in Table 4.5.

Construct	Kd for folded sample (µM)	Kd for refolded sample (µM)
WT	0.71 ± 0.11	0.84 ± 0.11
L26A	0.83 ± 0.27	0.88 ± 0.28
L254A	0.77 ± 0.31	0.88 ± 0.12
I34A	0.81 ± 0.23	0.74 ± 0.25
I258A	0.95 ± 0.37	0.71 ± 0.27
Y38A	0.86 ± 0.22	0.72 ± 0.29
Y266A	0.72 ± 0.22	0.80 ± 0.11
V43A	1.34 ± 0.21	1.19 ± 0.20
L267A	0.88 ± 0.24	0.86 ± 0.36
L44A	0.65 ± 0.26	0.73 ± 0.22
L268A	0.64 ± 0.30	0.70 ± 0.32
L50A	0.83 ± 0.21	0.70 ± 0.25
L273A	0.67 ± 0.16	0.80 ± 0.22
L54A	0.66 ± 0.21	0.83 ± 0.30
L279A	0.88 ± 0.28	0.85 ± 0.30

Table 4.5 Binding constants for folded and refolded samples of WT and alanine mutants

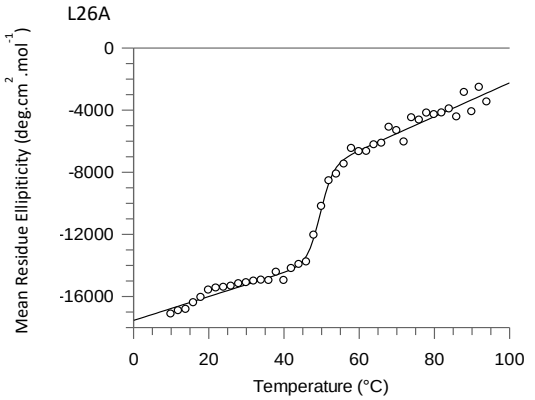
The binding constants from the saturation binding curves shown in figure 4.14 are listed here in order of mutations placed at the cytoplasmic side of the membrane to the mutations located at the side of the periplasm. The error are taken from the standard error of the fit.

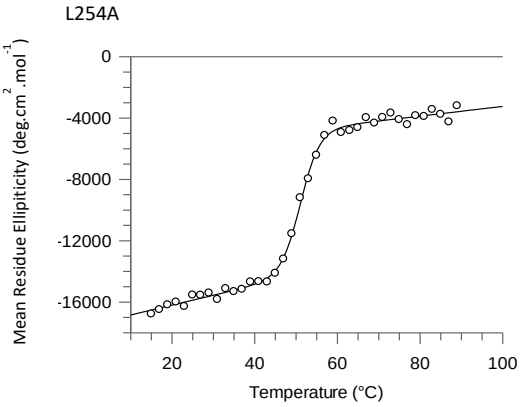
4.2.3 Temperature stability

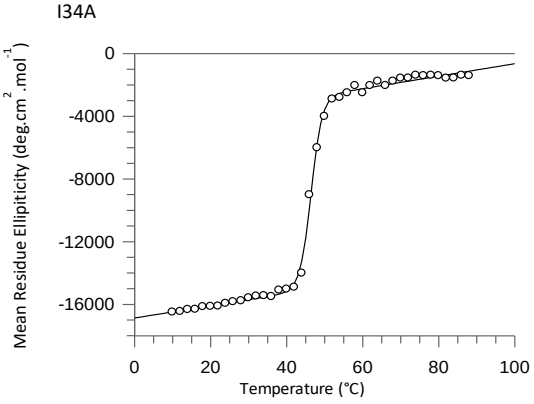
Thermal denaturation was used to provide an additional parameter to compare the effect of the alanine mutations on protein stability. Temperature melts are an easy and inexpensive method of comparison, as they require little protein to obtain a full set of data. In these experiments, the chosen protein is incubated at increasing temperatures, while the measurements of the current protein state are being taken. However,

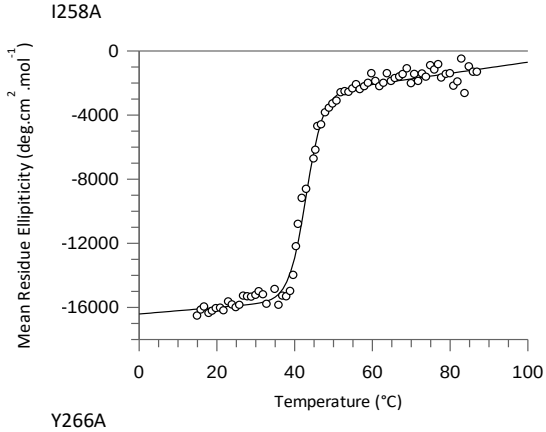
temperature denaturation data shown does not provide means of calculating ΔG values, as the temperature denaturation of GlpT, like in most membrane proteins, is irreversible. The change in GlpT structure incubated at each increasing temperature step is recorded by following the decay in the CD signal at 222 nm. The midpoint of the transition was calculated from the two state curve and recorded in denaturation data set, shown in Table 4.6. Figure 4.15 shows individual denaturation curves for each alanine mutant. For comparison, the graphs are positioned with mutations on helix 1 located on the left and mutation on helix 7 on the right side of the figure.

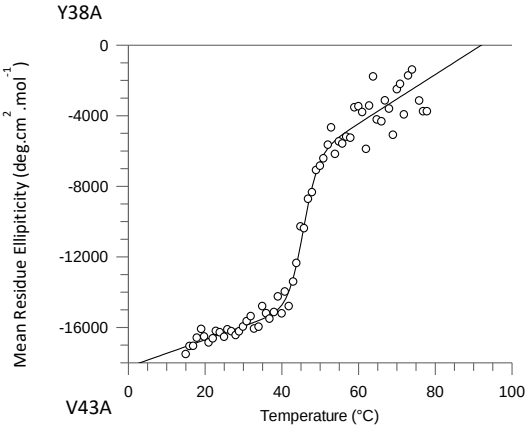
Overall, there is little difference in changes of thermal stability of alanine mutants. There is a higher frequency of helix 1 mutations decreasing the Tm of GlpT. The most disruptive mutations were Y38A and L44A. The mutations that showed the most stabilising effects were L279A, L273A, and L267A. The mutant pair Y38A-Y266A showed smaller Tm, compared to the WT and other alanine mutants. The change could be a result of increased cavity caused by an alanine substitution in the area of GlpT's binding pocket, or the elimination of tyrosine OH group as a hydrogen bond acceptor. Alanine mutations located on helix 7 of GlpT, seem to increase temperature stability of the protein. Interestingly, alanine mutant urea denaturation experiments have shown two helix 7 mutations (L268A and L273A) as strongly stabilising, compared to only one mutation on helix 1 (L54A). However, any proposition based on the temperature denaturation data should be taken carefully, as the method of denaturation itself is irreversible and the data is susceptible to noise. Care was taken to standardise each thermal denaturation experiment by equal protein concentrations and buffer conditions as the likelihood of protein aggregation and presence of multimeric species have the potential of skewing these results.

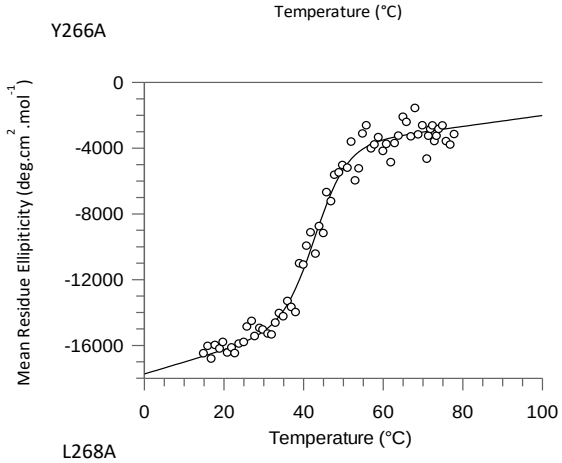

L26A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)

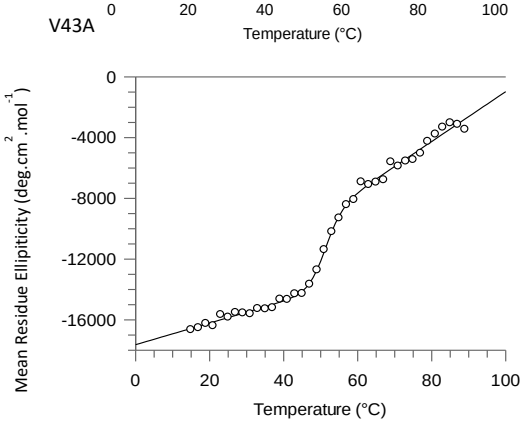

L254A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
20
40
60
80
100
Temperature (°C)

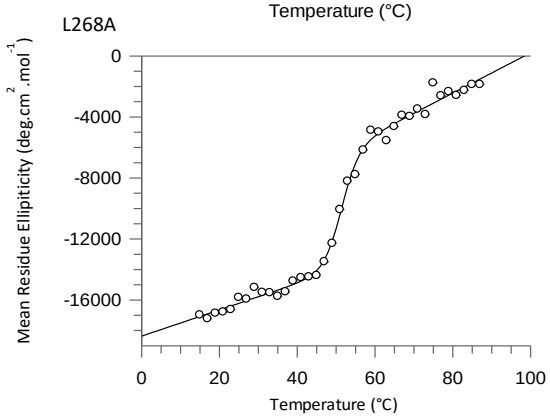

I34A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)


I258A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)


Y38A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)


Y266A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)


V43A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)


L268A
Mean Residue Ellipiticity (deg.cm^2.mol^{-1})
0
-4000
-8000
-12000
-16000
0
20
40
60
80
100
Temperature (°C)

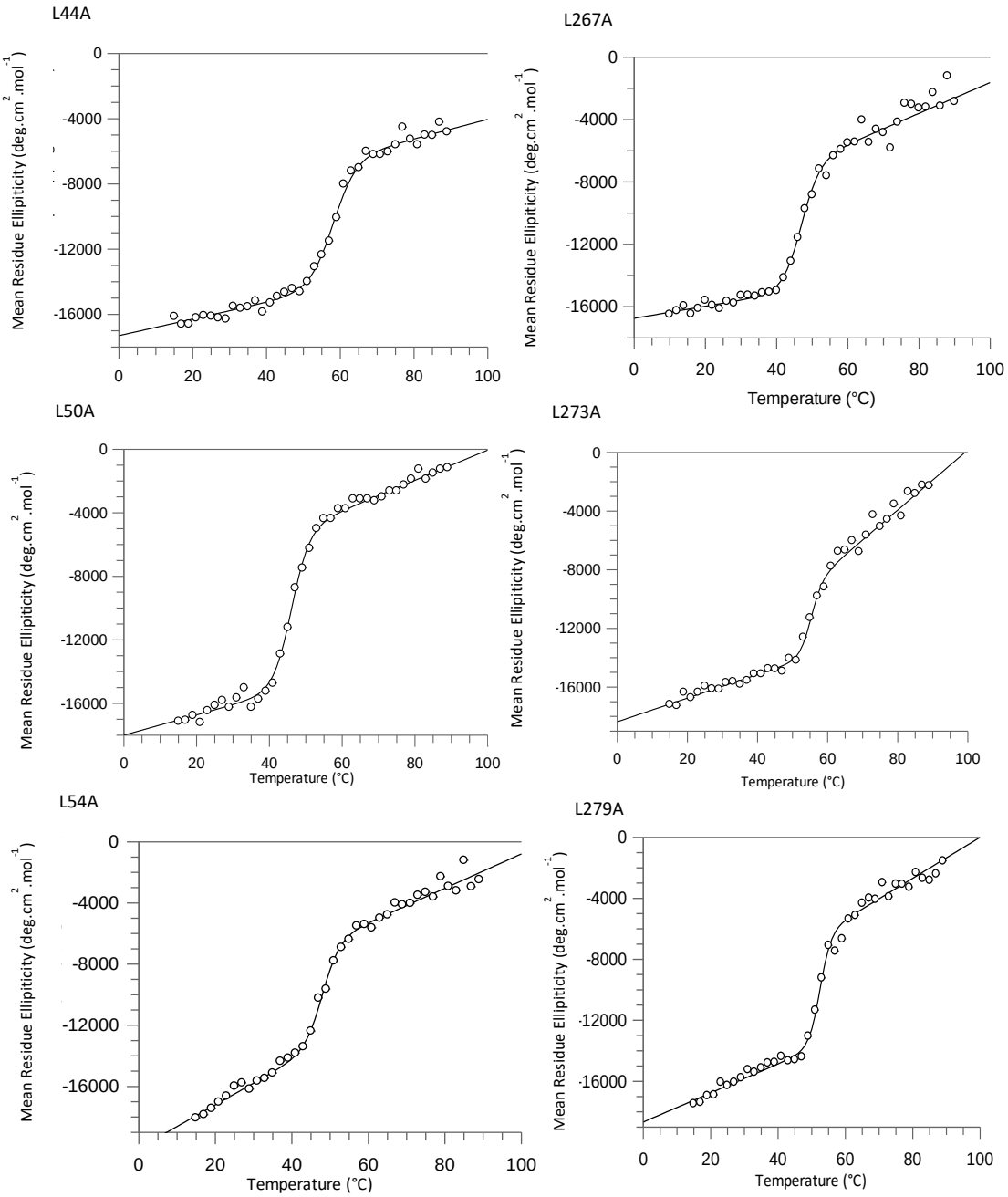


Figure 4.15 Thermal denaturation of alanine mutants

The changes in structure were monitored by measuring CD signal at 222 nm band between temperatures of 15 °C to 90 °C at 2 °C intervals. Each protein melt was achieved using 0.1 mg.ml^{-1} concentration of the protein in a 0.2 mm pathlength cell at 1 nm intervals and 1 second acquisition time from 260 nm to 180 nm in buffer. The melt curves were fitted with sigmoidal plot to calculate the midpoint on the slope of transition. The comparison graph of different Tm values obtained is found in figure 4.15 and summary in table 4.6. The raw data was plotted, fitted and converted to MRE in GraFit software.

Mutation	Tm (°C)	ΔTm
WT	49.0 ± 0.2	
L26A	49.9 ± 0.6	-0.9
L254	51.5 ± 0.3	-2.6
I34A	48.4 ± 0.1	-0.5
I258A	48.9 ± 1.4	-0.1
Y38A	45.5 ± 0.1	3.4
Y266A	42.5 ± 0.2	5.5
V43A	50.6 ± 0.3	-1.7
L268A	51.7 ± 0.5	-2.8
L44A	46.8 ± 0.4	2.2
L267A	54.5 ± 0.9	-5.6
L50A	48.9 ± 0.3	0.1
L273A	55.0 ± 0.5	-6.1
L54A	48.1 ± 0.4	0.9
L279A	52.4 ± 0.4	-3.5

Table 4.6 Summary of the temperature denaturation values from alanine mutants

The Tm values were taken from the denaturation scans, shown in figure 4.14. Each melting midpoint was calculated by fitting sigmoidal equation to denaturation curve for each alanine mutant, the errors are given from the standard error of the fit. The difference between individual mutants and a wild type was calculated as a $\Delta Tm = Tm_{(WT)} - Tm_{(mut)}$.

4.3 Discussion

4.3.1 Two state folding model for alanine mutants

The two-state model is the simplest way of analysing the unfolding and refolding data, as only folded and unfolded states are taken into consideration. The unfolding and refolding curves show only two phases with one transition phase around 5 M urea midpoint. However, the unfolding plots are made out of data points collected at 1 M urea intervals. The smaller data intervals may uncover smaller state within the transition phase of the unfolding that could have been missed otherwise.

Further experiments should include time-resolved measurements of folding and refolding. Similar to the measurements performed at steady-state, the rates of folding and refolding should be in agreement with each other, showing the same linear dependence on the denaturant concentration. Such results would further confirm the two-state model for GlpT, as well as show the existence of any kinetic intermediate in both folding and refolding assay. Moreover, the kinetic measurements would present an alternative method of measuring stability, and combined with the steady state measurements identify the position of the transition state through the alanine mutation analysis in the helices 1 and 7.

Attempts at measuring kinetics of GlpT were described in chapter 3. It is possible to record the unfolding rate of GlpT WT, however the rich aromatic environment of GlpT complicates the fitting of the data to a single exponential fit. A potential improvement to this experiment would be substitution of many intrinsic fluorescence reporters with a single probe substitute. The single reporter could be used to measure kinetics of the unfolding and thus gain insight into the state of the protein during the unfolding. Design of a rhodamine marker thought to characterize behaviour of GlpT during unfolding is further discussed in chapter 3.

The urea-induced reversibility was shown for all alanine mutants through the regain of CD signal measured at 222 nm, as well as the return of the binding constant. Although there was a small discrepancy at a few urea concentration points, the general trend of the refolding curve suggests that the refolding pathway of GlpT's alanine mutants is similar to the unfolding pathway and the alanine mutations do not seem to affect the pathway. Further repeats of the points on the refolding curve would provide greater resolution into the refolding curve and confirm the close relationship between unfolding and refolding curves.

The comparison of ΔG_U^{H2O} and ΔG_f^{H2O} calculated from the unfolding and refolding data shows similarities between these values. Both the ΔGu^{H2O} and ΔGf^{H2O} were found to be within the region of 4.4 kcal.mol^{-1} and therefore are an indication of the unfolding reaction being in a reversible equilibrium. Furthermore, the similarity between the slope of the unfolding curve, m_U^{H20}, and slope of the folding curve, m_F^{H20}, as well as the midpoint Cm values suggest that the protein is in equilibrium, changing its state between the folded and unfolded phases. The refolding assay, measured by CD spectroscopy, has a higher error associated with the calculation of m_F^{H20} and Cm values than the calculation of unfolding's m_U^{H20} and Cm values, resulting in an uncertain final value of ΔGf^{H2O}, especially for the mutants L26A and V43A.The reason for such discrepancies could be due to the refolding efficiency of mutants being reduced, compared to the WT, thus contributing to the larger spread of the refolding values in the repeated experiments. Also, the discrepancy in V43A refolding could be linked to its lower stability measured by the unfolding assay.

4.3.2 Differences in stability of alanine mutants

The alanine substitutions along the helix 1 of GlpT were designed to investigate the possibility of a stable unit located in the first helix of the membrane protein. As a comparison, a residue in a similar environment and position was also mutated to alanine

in the opposite helix 7. Both helices are the first helices of their respective domains and therefore any difference to the stability of helix 1 of the protein could be attributed to its location as the first helix to be translated by the ribosome.

A series of alanine mutants were investigated using urea as a chemical denaturant in the unfolding assay. Although, as shown in chapter 3, a stronger denaturant GuHCl unfolds up to 45 % of the total protein structure, its use was omitted due to lack of refolding upon dilution of the denaturant. Overall, it was found that helix 7 mutants unfold less in the 8 M urea compared to the WT. Helix 7 mutations showed an average 11% decrease in the total MRE reduction, compared to helix 1 mutations. Helix 1 mutations showed a 6% increase in the total MRE reduction. Only the V43A mutation showed less unfolding. These results suggest that alanine substitutions in the helix 1 lead to an increased degree of unfolding in urea within the structure of GlpT. However, a number of mutations that lead to a total MRE change upon unfolding is higher in helix 7. 6 helix 7 mutations had a significant effect on total MRE change, compared to 4 helix 1 mutations. The helix 7 has therefore shown to be more sensitive to mutations than helix 1.

The protocol of GlpT WT reversibility, first established in chapter 3, was used to assess whether the alanine substitutions have any influence on the capacity of the protein to refold from the urea denatured state. The GlpT unfolding curves were fit to a two-state curve and the ΔGu^{H2O} values were calculated. The extrapolation of data from the linear fit introduces a larger error into the final analysis. Therefore the calculated values from the sigmoidal fits were used to form the final comparison.

The measured effect of alanine substitutions on the GlpT ΔG_U^{H2O} values range from 3.9 ± 0.6 to 4.6 ± 0.6 kcal.mol^{-1}. Large errors and spread of data associated with the ΔG_U^{H2O} values need to be considered when reviewing these results. The spread of these values

is the result of measuring small differences in energy caused by the elimination of a few individual methyl groups across the length of the helix.

Overall, considering the large error, the majority of the mutations were found to have a general stabilising effect on helix 7, while 3 mutations on helix 1 had a destabilising effect compared to 1 destabilising mutation on helix 7. The changes in ΔG_U^{H2O} compared to the WT are both positive and negative, when comparing different mutations (table 4.3). However, when considering the mutation location within the protein structure, the mutations placed in the protein core were predominantly destabilising, with only one mutation on helix 7 being stabilising. By comparing the mean WT ΔG_U^{H2O} value against the largest difference in ΔG_U^{H2O} of alanine mutants (>0.20 kcal/mol), helix 1 mutations showed a higher frequency of destabilising mutations (V43A, L44A and L50A) compared to the helix 7 mutation (I258A). The higher occurrence of helix 1 $\Delta\Delta Gu^{H2O}$ WT destabilising mutations shows the possibility that helix 1 may be slightly more sensitive to alanine substitution than helix 7. The comparison between the differences in energy within mutation pairs showed little or no difference in the mutations placed on the cytoplasmic side of the membrane, such as the L26A-L254A pair, as well as near the binding pocket, the Y38A-Y266A pair. Interestingly, the mutations Y38A, Y266A, V43A, L44A, L267A and L50A, located in the protein core and facing towards other helices, showed a repeated destabilising effect on the ΔG_U^{H2O} value. The next mutation pair L54A-L279A, located at the very end of the helix, contradicts this statement. This final mutation is located at the very end of the helix on the periplasmic side of the membrane and could therefore be counted as being influenced by the DDM molecules forming a micelle around the protein. A summary of the ΔGu^{H2O} values from alanine substitutions can be found in table 4.7 and presented on a diagram in figure 4.16.

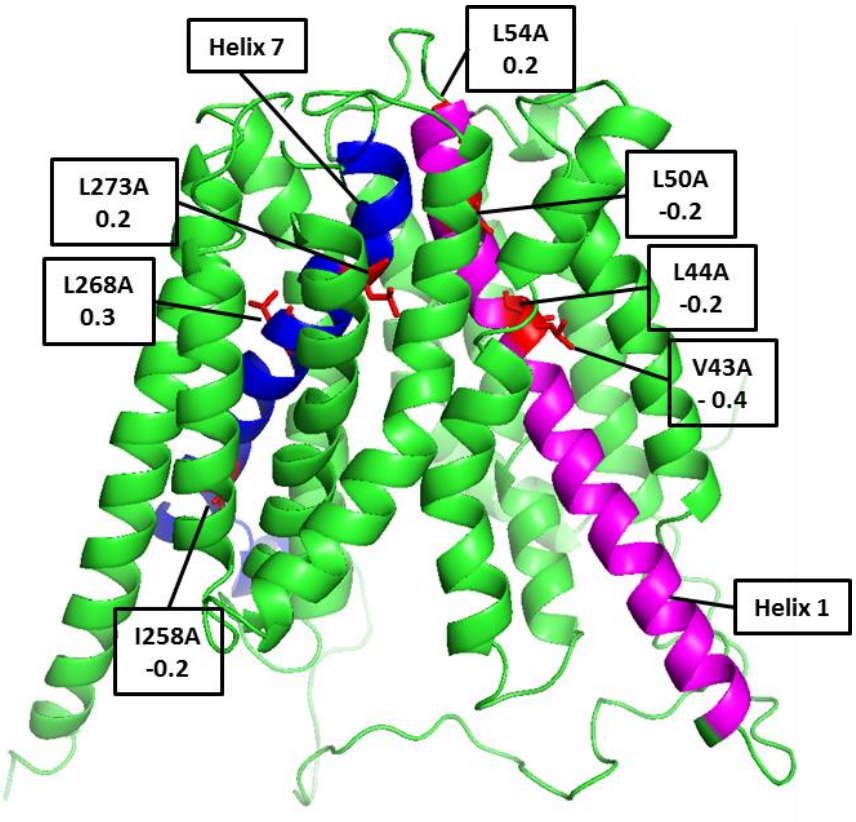


Figure 4.16 Diagram of alanine substitutions and their effect on free energy of unfolding in the helices 1 and 7

Mutation sites that had an effect on free energy of unfolding are presented above with their respective $\Delta\Delta Gu^{H2O}$ WT, calculated as a ΔGu^{H2O} WT - ΔGu^{H2O} Ala shown below. Helix 1 is indicated by purple colour and helix 7 by blue colour. The summary of all of ΔGu^{H2O} values from Ala mutants are shown in table 6.1. Structure was rendered in PyMOL software.

Furthermore, the mutations that were located in the vicinity of the protein core have been found to be predominately destabilising. Possibly due to alanine substitutions creating cavities that increase the access of the denaturants, making the protein core more susceptible to the effects of denaturation. Alternatively, the binding of P5P increases resistance to denaturation, possibly *by* reducing the access of the denaturants to the binding site.

Construct	ΔGu^{H2O} (kcal.mol^{-1})	$\Delta\Delta Gu^{H2O}$ WT	Effect
WT	4.3 ± 0.5		
L26A	4.2 ± 0.4	-0.1	Minimally altered
L254A	4.2 ± 0.8	-0.1	Minimally altered
I34A	4.4 ± 0.4	0.1	Minimally altered
I258A	4.1 ± 0.2	-0.2	Destabilising helix 7
Y38A	4.3 ± 0.3	-0.1	Minimally altered
Y266A	4.2 ± 0.7	-0.1	Minimally altered
V43A	3.9 ± 0.5	-0.4	Destabilising helix 1
L268A	4.6 ± 0.5	0.3	Stabilising helix 7
L44A	4.1 ± 0.4	-0.2	Destabilising helix 1
L267	4.2 ± 0.4	-0.1	Minimally altered
L50A	4.1 ± 0.3	-0.2	Destabilising helix 1
L273A	4.5 ± 0.9	0.2	Stabilising helix 7
L54A	4.5 ± 0.5	0.2	Stabilising helix 1
L279A	4.2 ± 0.8	0.1	Minimally altered

Table 4.7 Summary of ΔGu^{H2O} values from Ala mutants from two state curves

The free energy values for alanine mutants and WT were calculated from sigmoidal fits to CD unfolding data, errors are taken from the standard error of the fit. The $\Delta\Delta Gu^{H2O}$ was calculated as a ΔGu^{H2O} WT - ΔGu^{H2O} Ala difference.

A higher frequency of destabilising mutations near compact protein centre can be explained using the solvation model. Cavity-creating mutations, such as alanine substitutions increase the access of the denaturants, making the protein core more susceptible to the effects of denaturation. Moreover, these mutations could have also

disrupted weak intermolecular interactions between helices, thus affecting the packing energetics and efficient burial of hydrophobic residues (181).

It is difficult to draw a definitive conclusion due to the error of the ΔG_U^{H2O} and the complex nature of intermolecular interactions within the protein structure. The errors associated with the results make it difficult to fit the data accurately and subsequent data manipulations carry a bigger error. As the contribution from different mutation environments was reduced, it should be easier to define the differences in stability between the helices. A potential improvement to distinguish between the stability of helix 1 and 7 would be to either use amino acid mutation that is larger than alanine, such as valine or isoleucine, or use different method of reporting the unfolding of the protein. Furthermore, additional controls which increase protein preparation reproducibility would be advantageous in decreasing error seen in CD measurements. Fluorescence can be measured very accurately, however it may not necessarily report the loss of structure. A combination of fluorescence tags with other methods of measuring the structure change, could be used to collect data to supplement the study. The individual tags could be used to report on the extent of unfolding in the labelled regions of the protein.

4.3.3 Temperature induced denaturation

The alanine mutants were subjected to the additional stability assay *via* means of temperature-induced denaturation. Overall, an increase in the thermal stability was observed for the mutations located in helix 7, with the exception of the Y38A-Y266A mutation pair. Mutations in helix 1 showed marginal or no difference in stability when compared against the WT. Although the increase in thermal stability (table 4.6), caused by mutations in helix 7, overlaid with the results from the chemical denaturation assay (table 4.1), they are not necessarily linked, as the two assays involve different mechanisms of unfolding, resulting in two different pathways of obtaining the unfolded state. However, the temperature denaturation data provides an additional insight into the effects of alanine substitutions. Most mutant Tm values fall within 3.5 °C difference from

the Tm measured for the WT, with the exception of Y266A, L267A and L273A, which is an indication that overall the alanine substitutions do not greatly affect the general stability of the protein to the temperature denaturation. The Y266A mutant showed a flatter gradient compared to the rest of the alanine mutants. The result could indicate that the sample of the protein used has a more heterogeneous population of protein species, therefore affecting the slope of the transition phase. Overall, the alanine substitution mutations did not have big enough effect to change the behaviour of GlpT during temperature-induced denaturation.

4.3.4 P5P binding of alanine mutants in the folded and refolded state

Reversibility of the unfolding reaction was tested *via* a binding assay, in addition to a simpler refolding assay by a method of rapid dilution of urea. The ligand P5P was picked as the best ligand to test the return of the binding affinity to GlpT. The binding of P5P was followed by the measurement of the CD signal at 222 nm. The large dilution required during the process of refolding did not impede data collection, as the P5P binding was strong in a magnitude of 0.7 µM, requiring little protein to successfully collect a complete saturation binding curve and to avoid protein aggregation issues during the experiment. All alanine mutants have shown successful recovery of binding affinity, with the average Kd for folded protein equal to 0.81 µM, compared to the 0.82 µM for refolded protein, showing return of function. The error in the CD measurements could be attributed to the slight differences in the protein concentration between the samples, as the unaccounted protein concentration would result in a CD signal difference imitating binding of the P5P. Alternatively, binding of P5P ligand, discussed in chapter 3, had a bigger effect on temperature and chemical stability of GlpT WT.

The equal recovery of the CD signal, measured in the refolding assay, and recovery of the P5P ligand affinity, measured in the binding assay, suggests that both secondary and tertiary structure return to their native state upon GlpT refolding. Investigation into

the recovery of binding of different ligands bound in other areas of protein would further confirm the return of inter-helical interactions making up the tertiary structure.

The substrate assay provides information on the recovery of binding affinity during protein refolding, but also informs the effect of alanine mutations on the potential changes in the binding pocket affecting that binding affinity. Surprisingly, Y38A and Y266A mutations did not significantly affect P5P binding, confirming the claim from the previous studies (119) in which these residues are not directly involved in binding, but rather act to close the channel on one side when the ligand is transported to the other in the rocker-switch-type transport mechanism. A small loss of affinity was noted for the V43A mutation. It is unclear why an alanine substitution would affect the binding affinity within the protein structure. The residue is located away from the binding pocket and is facing towards small hydrophobic residues located on helices 3 and 6. However, it is possible that the mutation location affected the protein interaction with the detergent molecules within the DDM micelle, in turn causing a minor change in conformation of helices and therefore a drop of the binding sensitivity. Further investigation into the binding and transportation mechanism of GlpT would clarify the issue and prove or disprove the role of the V43 residue. An assay involving G3P-Pi transportation across the membrane could be used to show the return of function upon GlpT refolding as well as the effect of the V43 on the P5P mediated inhibition of G3P transportation across the membrane.

4.4 Summary of chapter

- Alanine mutants show recovery of structure and the return of P5P binding during the refolding assay.
- The effect of alanine mutations on P5P binding shows the lack of Y38 and Y266 involvement in binding of P5P.
- Alanine mutants display a range of ΔGu^{H2O} values with mutations located in the protein core showing a decrease in stability.

Chapter 5

Final discussion and future work

5.1 Aims of the chapter

This section is summarising the work in this thesis, highlighting key outcomes and discussing their implications, as well as possible improvements in measuring inter-helical stability. At the end of this chapter, potential future work is discussed that would expand understanding of membrane protein folding.

5.2 Summary of results

The denaturation assays with urea and GuHCl have shown that GlpT's structure is reduced by approximately 33% and 45%, respectively. The urea unfolded state of GlpT has also been shown to be reversible with a complete return of the signal, as measured by far-UV CD spectroscopy, upon exclusion of urea. Although the harsher denaturant, GuHCl resulted in a greater reduction in of CD signal, it was not possible to recover the native structure upon GuHCl removal.

The unfolding of GlpT with urea denaturant is consistent with a two state reaction with only folded and unfolded states present in the unfolding reaction. Moreover, the GlpT was shown to fully regain both function and structure after refolding from its denatured state. Unfolding curves from both fluorescence and CD measurements estimate the ΔGu^{H2O} for the GlpT unfolding to be approximately 4.6 and 4.3 kcal.mol^{-1}, respectively. Regardless of the method of following the state of the protein, both ΔGu^{H2O} from unfolding and refolding reactions are similar in magnitude, supporting the equilibrium state reaction model of GlpT folding. Further evidence of the reversibility of urea denaturation was provided by the P5P binding assay. The results of the sample before denaturation match closely with the sample after refolding reaction. This suggests that the GlpT is able to regain its structure as well as tertiary conformation, as the amino acid residues responsible for binding have likely returned to their native positions before denaturation, thus resulting in a similar affinity to substrate binding.

The alanine substitutions of amino acid residues along the length of the helices 1 and 7 of GlpT have been designed to investigate differences in the stability of the two targeted helices. All Ala mutants refolded reversibly, similarly to WT. Only the V43A mutation was shown to affect the affinity of the chosen substrate, P5P, to its binding site. Out of 14 alanine substitutions, four (I258A, V43A, L44A and L50A) have shown a potential decrease in ΔGu^{H2O} values, with the remaining alanine substitutions showing a range of values. V43A has showed to be most destabilising, possibly due to the effect on helical packing of the surrounding helices 3, 4 and 6, thus increasing the accessibility of the denaturant. The alanine scan has also shown a higher likelihood of destabilising mutations near the protein core and facing other helices, compared to the mutation sites facing the protein surface, as these are less likely to cause a destabilising effect. These results highlight the effect of cavity-creating mutations on access of denaturant as well as destabilising effect of disrupting weak intermolecular interactions between helices, thus affecting the packing energetics and efficient burial of hydrophobic residues (181).

5.3 Discussion

Overall, the GlpT unfolding with urea shows approximately 33 % of protein structure loss. Taking into account the resistance of the helical structure to denaturation, it is likely that the exposed regions that resemble the unfolded state are aqueous loops, ends of the α-helices as well as the binding cavity separating the two domains. The remaining structure remains mostly folded with only most exposed regions being affected by the denaturation. Interestingly, binding of the P5P substrate results in the marked increase in resistance to denaturation, possibly due to reduced accessibility of the denaturant to the binding site. Moreover, the initial results with the rhodamine conjugated GlpT suggest that the unfolding of the unbound protein starts at the interface region; same region is also responsible for the binding of the substrate. The role of the binding cavity in facilitating of unfolding may be due to three components:

- the critical position at the interface of two domains;

- the specific chemical composition of hydrophilic amino acids to facilitate transport;
- the enabling of access to the denaturants.

Out of all the alanine substitution mutations, the V43A-L268A pair showed the greatest difference in stability compared to the WT. The pair is located at the interhelical interface between the helices 3, 4, 6 and 9, 10, 12 for V43A and L268A, respectively. The explanation why these mutations caused a much bigger difference compared to other mutations has yet to be resolved, but it could be result of altering interactions and packing between the neighbouring helices. Overall, majority of the mutations were found to have a general stabilising effect on helix 7, while 3 mutations on helix 1 had destabilising effect compared to 1 destabilising mutation on helix 7.

Furthermore, the mutations that were located in the vicinity of the protein core have been found to be predominately destabilising. This is possibly due to alanine substitutions creating cavities that increase the access of the denaturants, making the protein core more susceptible to the effects of denaturation. Alternatively, binding of P5P increases resistance to denaturation, possibly *via* reducing the access of the denaturants to the binding site.

The differences in behaviour of mutations in helices 1 and 7 were noted. Although it is difficult to say how these can be compared to the WT, they did show a range and variety of ΔGu^{H2O} values, regardless of their possible gene duplication evolutionary history. These findings are in agreement with the previous work on LacY, GalP and bR, where the protein unfolds mostly in areas of the hydrophilic cavity or domain interface and different areas of the protein may have significantly different stability *in vitro*. These shared attributes could mean that a set of rules governs the protein folding in vitro and could be used in future studies on therapeutically relevant targets (162). Moreover, the results presented in this thesis show a complex range of amino acid contributions to stability of helices in two domains, regardless of two domains pseudo symmetry.

Although the alanine scan analysis consists of large errors from CD measurements and protein instability, it is a useful inquiry into the behaviour of membrane transporters as there are not many folding studies for large multi subunit transporters focusing on the relative stability of its components. The development of the rhodamine labelled GlpT shows promise and proves its ability to report on fluorescence changes during the unfolding without it causing any change to the CD spectra as measured by far-UV CD. It could be used therefore, as a reporter of protein changes during the unfolding studies. However, the question whether a larger label is influencing protein conformational changes during the experimental processes still needs further investigating.

5.4 Comparison with other MFS protein folding studies

Similarly to other alpha helical membrane proteins such as XylE, GlpT shows a high resistance to unfolding by thermal and chaotropic denaturation. The total amount of CD signal lost due upon urea denaturation is comparable to studies of MFS transporters solubilised in DDM, LacY and GalP. All three MFS transporters remained folded in 2 M urea, unfolding starting around 3 M concentration of urea, with midpoint of unfolding curve at around 4 M concentrations. Furthermore, GlpT's full return to the native structure after urea-induced unfolding has also been noted in the reversible denaturation reactions with other members of the MFS such as, LacY, GalP and XylE.

The determined ΔG_U^{H2O} for GlpT is estimated to 4.3 ± 0.5 kcal.mol^{-1} and is within the same magnitude range as its related proteins LacY with the ΔG_U^{H2O} of 2.2 ± 0.9 kcal.mol^{-1}. The higher stability of the GlpT is reflected by its larger size of 52 kDa compared to LacY's 46 kDa. However, GalP with molecular mass, estimated to be 50 kDa, was determined to unfold with ΔG_U^{H2O} equal 2.5 kcal.mol^{-1}, compared to GlpT's 4.3 kcal.mol^{-1}. Unlike GalP and LacY, GlpT has a higher energy of unfolding per residue of 0.10 kcal.mol^{-1} per residue, similar to DGK (0.13 kcal.mol^{-1}) and bR (0.08 kcal.mol^{-1}).

The lower unfolding energy values for these MFS transporters compared to GlpT could possibly be explained by their proposed mode of transport, differences in flexibility and therefore access of denaturant to inter-domain site, rather than differences in size. GlpT is an example of antiporter, compared to the LacY's and GalP's symporter function. The nature of the substrate transporter could have effect on the mechanism of transport and flexibility of the structure that accommodates the protein's function. The differences in the domain stability and the importance of the binding cavity as a domain interface have been reported as key characteristics of the MFS protein stability in the previous LacY work (106). Interestingly, an increase in stability to chemical unfolding has only been reported in GlpT and XylE (in process of publication). Both LacY and GalP did not show an increased stability when bound to their respective substrates.

Although GlpT shares common fold with GalP and LacY, results presented in this thesis also highlight some differences in the behaviour and stability, which need to be addressed by further studies on MFS transporters in order to form a set of rules governing the stability and folding of these group of membrane proteins. The importance of inter-domain interface in mediating interactions and unfolding with denaturant is seen in both LacY and GlpT studies. Moreover, studies of these two-domain proteins have both highlighted that, although the MFS domains are pseudo-symmetrical, they do not share same stability in-vitro. Furthermore, differences in stability between members of the MFS family show a complex picture of membrane protein folding. Nevertheless, the detail obtained from these studies puts forward the methodology needed to establish future studies focused on physiologically this important family of transporters.

5.5 Focus of the future studies

Although, the alanine mutation analysis is incomplete and it did not show clear differences between two helices or domains, it still provided the methodology, as well as highlighted potential difficulties and pitfalls that will be useful for the development of future studies. This work shows the potential of multi technique experiments in deducing

the behaviour of membrane proteins.

The potential improvements could involve the use of less conservative, bulkier mutations, rather than alanine substitutions, such as valine or isoleucine, to cause bigger, more distinguishable changes to the stability. The bigger differences in stability could be easier to compare against the error associated with the measurements. Moreover, the combination of fluorescence tags with other methods of measuring structure change, could be used to supplement and collect more informative data. The individual tags could be used to report on the extent of unfolding in the labelled regions of the protein. Chapter 3 provides the groundwork to expand the alanine scan of helices methodology to include fluorescent tags. Once, the system of labelling is optimised and fully characterised it could be used to measure time-resolved unfolding data. Furthermore, by placing two rhodamine tags at different positions, it would be possible to determine the relative stability and behaviour of those protein regions compared to the overall picture of GlpT's unfolding. For example, one of the possible studies would be to characterise the flexibility and stability of periplasmic and cytoplasmic regions of GlpT during the process of unfolding. The same system could also be used as a sensor of substrate binding, as GlpT undergoes extensive conformational changes during the process of binding.

The established two state unfolding system of GlpT in urea in micelles could be modified to extend the GlpT's unfolding studies to include other membrane mimetic systems, such as liposomes. The native membrane environment differs significantly from the experimental *in vitro* conditions. First, the two state and the reversible equilibrium of unfolding need to be shown for liposomes. Then, some of the investigated mutants could be tested in the lipid system that is more related to the *in vivo* conditions. A different hydrophobic environment may provide a bigger contrast to the free energy comparisons between the pairs of mutants. The combination of fluorescence tags and liposomes could be used to see how GlpT's flexibility and stability are affected by the bilayer.

One of the key features of the MFS transporters is their interface region, found between the two domains. The interface is made up of residues creating a hydrogen bond network through the centre of the protein, which restricts the movement of molecules through the transporter depending on whether the recognised substrate is bound. The interface is also lined with a few charged amino acids that facilitate the binding to the transported substrate. Since, the interface region is a key characteristic of the MFS transporters, studying its contribution towards the stability of the protein could expand its role in protein folding events and its role in the mechanism of substrate transport. This could be achieved by selective mutations disrupting the key interactions between the two domains, coupled with FRET or H\D exchange mass spectroscopy.

Appendix A. Primer sequences

Primers used in mutagenesis reactions that led to experiments in chapters 4 and 5 are listed below. The ‘fwd’ suffix is used to describe primers directed for the sense, forward strand, while primers with ‘rev’ suffix were used to bind to the reverse, antisense sites. The matching nucleotides are written as small case, mismatching sequences are represented by capital letters. The primer sequence is supplemented with the calculated melting temperature (Tm) with and without the mutation site. Calculation of Tm and other oligonucleotide properties was performed using oligo property scan tool web service found on the primer’s supplier website (https://ecom.mwgdna.com/services/webgist/mops.tcl). Table lists primers used in alanine scan experiments, all mutations with the exception of L26A, L268A and L44A were performed with regular site-directed mutagenesis protocol. Remaining primers were used in the ‘Round the Horn protocol.

Mutation	Sequence (5' → 3')	Tm (°C)	Tm without mutation (°C)
Alanine scan mutations			
L26AHornfwd	GCgcgctggcaaattttcctg	63.2	57.5
L26HornArev	tcgacgataagtcggatcgatctc	65.2	65.2
L254Afwd	gccgaacaaaGCgctgtggt	62.5	56.0
L254Arev	accacagcGCtttgttcggc	62.5	56.0
I34Afwd	cgctggcaaGCtttcctg	58.2	51.7
I34Arev	caggaaaGCttgccagcg	58.2	51.7
I258Afwd	tgctgtggtatGCcgccatc	61.4	51.0
I258Arev	gatggcgGCataccacagca	61.4	51.0
Y38Afwd	ctttggcGCTgcggcttactatttg	64.6	60.0
Y38Arev	caaatagtaagccgcAGCgccaaag	64.6	60.1
Y266Afwd	gtgttcgttGCTctgctgcgttac	64.4	55.8
Y266Arev	gtaacgcagcagAGCaacgaacac	64.4	55.8
V43AHornfwd	GCggttgagcagggat	52.8	48.6
V43AHornrev	ataaggcatagcaagcgc	51.6	51.6
L268AHornfwd	Cgcgttacggcatcctc	57.6	54.3
L268AHornrev	Ccagataaacgaacacgttgg	57.9	55.3
L44Afwd	gctatgccttatGCggttgagca	62.5	55.8
L44Arev	tgctcaaccGCataaggcatagc	62.5	55.8
L267Afwd	cgtgttcgtttatGCgctgcgtta	62.7	58.2
L267Arev	taacgcagcGCataaaacgaacacg	62.7	58.2
L50Afwd	gaactttgcgGCtgctatgccttatc	64.8	61.0
L50Arev	gataaggcatagcaGCcgcaaagttc	64.8	61.0
L273Afwd	cgttacggcatcGCcgactggt	67.2	60.4
L273Arev	accagtcgGCgatgccgtaacg	67.2	60.4
L54Afwd	gctatgccttatGCggttgagca	62.4	55.3
L54Arev	tgctcaaccGCataaggcatagc	62.4	55.3
L279Afwd	caccgacttatGCgaaagaggt	57.8	49.8
L279Arev	acctctttcGCataagtcggtg	57.8	49.8

Single-Trp mutations			
M107Wfrw	gcggcggcagtgTGGttgtttatg	66.1	62.8
M107Wrev	cataaacaaCCAcactgccgccgc	66.1	62.8
F148Wfrw	gtactatggtgcactGGttctcgcag	66.4	58.8
F148Wrev	ctgcgagaaCCagtgcaccatagtac	66.4	58.8
V332Wfrw	gcaaccggcTGGttctttatgac	62.4	54.8
V332Wrev	gtcataaagaaCCAgccggttgc	62.4	54.8
Rhodamine label construct mutations			
C215Sfrw	cgcaatcctCAggcttgc	58.2	54.3
C215Srev	gcaagccTGaggattgcg	58.2	54.3
S61Cfwd	gcagggattctGTcgcggtgat	64.0	61.4
S61Crev	atcaccgcgACagaatccctgc	63.4	61.3
V283Cfwd	ctgaaagagTGtaagcatttcgcgcta	63.4	61.3
V283Crev	tagcgcgaaatgcttaCActctttcag	63.4	61.3

Table 5 List of all mutagenesis primers used in creating tryptophan, cysteine and alanine construct. Mutations sites are indicated by capital letters

Appendix B. Summary of PCR conditions

The full protocols for Quik change mutagenesis and 'Round the horn are found in chapter 2 Materials and Methods sections. The optimal conditions for successful PCR amplifications for individual mutations are listed below.

Mutation	Tm without mutation (°C)	Optimal conditions for PCR amplification
L26A	57.5	2 mM Mg^{2+}, 55.5 °C for 30 cycles
L254A	56.0	3 mM Mg^{2+}, 51.0 °C for 35 cycles
I34A	51.7	3 mM Mg^{2+}, 47.7 °C for 30 cycles
I258A	51.0	4 mM Mg^{2+}, 57.0 °C for 30 cycles
Y38A	60.0	1 mM Mg^{2+},52.0 °C for 5 cycles, 54.0 °C for 5 cycles, 56.0 °C for 5 cycles, 58.0 °C for 5 cycles, 60.0 °C for 15 cycles, 2% DMSO
Y266A	55.8	2 mM Mg^{2+}, 50.8 °C for 25 cycles
V43A	48.6	2 mM Mg^{2+}, 44.6 °C for 35 cycles
L268A	54.3	2 mM Mg^{2+}, 49.8 °C for 25 cycles
L44A	55.8	2 mM Mg^{2+}, 50.8 °C for 25 cycles
L267A	58.2	2 mM Mg^{2+}, 53.2 °C for 25 cycles
L50A	61.0	2 mM Mg^{2+},53.0 °C for 5 cycles, 55.0 °C for 5 cycles, 57.0 °C for 5 cycles, 59.0 °C for 5 cycles, 61.0 °C for 15 cycles, 2% DMSO
L273A	60.4	2 mM Mg^{2+},52.4 °C for 5 cycles, 54.4 °C for 5 cycles, 56.4 °C for 5 cycles, 58.4 °C for 5 cycles, 60.4 °C for 15 cycles, 2% DMSO
L54A	55.3	3 mM Mg^{2+}, 50.3 °C for 25 cycles
L279A	49.8	4 mM Mg^{2+}, 45.8 °C for 35 cycles
M107W	62.8	3 mM Mg^{2+}, 57.8 °C for 30 cycles
F148W	58.8	2 mM Mg^{2+}, 53.8 °C for 30 cycles
V332W	54.8	1 mM Mg^{2+}, 49.8 °C for 30 cycles, 2% DMSO
C215S	54.3	3 mM Mg^{2+}, 49.3 °C for 25 cycles, 2% DMSO
S61C	61.4	2 mM Mg^{2+}, 46.4 °C for 25 cycles, 2% DMSO
V283C	61.3	2 mM Mg^{2+}, 46.3 °C for 25 cycles, 2% DMSO

Table 6 List of optimal conditions for each mutagenesis reaction performed

Appendix C. Growth curve for individual mutants

The LB growth media was established to provide highest yield of mutant of interest. Terrific broth led to expanded growth and increased bug weight, however it did lead to increased protein yield. As part of the growth protocol, media was preheated to 37 °C before overnight culture was added. Without the preheating, the lag phase of bacterial growth was increased by an additional 30 minutes. The descriptions of mutants and their final protein yields, are described below each growth curve. The yields for the gel purified proteins are given after spin concentrating approximately 8 ml of the peak elution of the gel filtration column, values are shown as mg of protein purified per 1 L bug culture grown, and the concentration was calculated by the UV absorbance at 280 nm. For the single tryptophan mutants, Markwell-Lowry assay was used to establish the total yield of the purification.

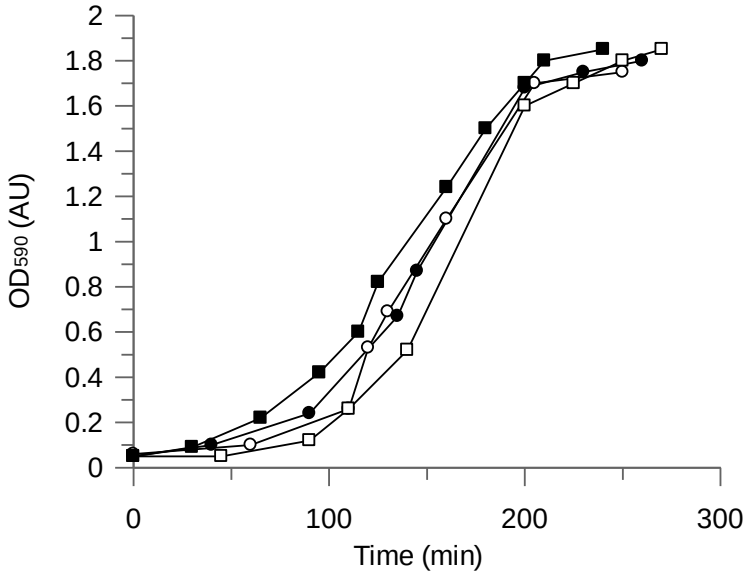


Figure 5 Growth curves for L26A, L254 and I34A alanine mutants

GlpT WT (open squares) yield of 1.02 mg/L, L26A (filled squares) yield of 1.05 mg/L, L254A (open circles) yield of 0.98 mg/L, I34A (filled circles), yield of 0.95 mg/L

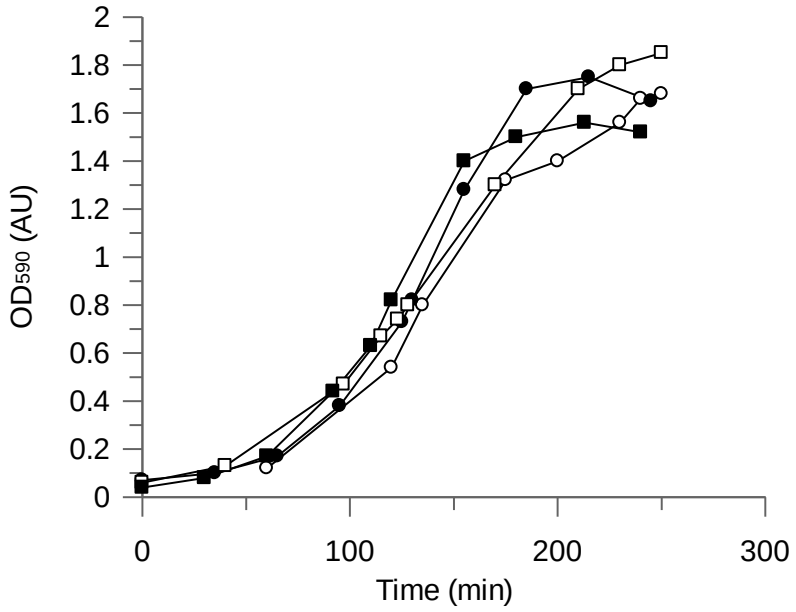


Figure 6 Growth curves for I258A, Y38A and Y266A and V43A alanine mutants

I258A (open squares) yield of 1.05 mg/L, Y38A (filled squares) yield of 1.01 mg/L, Y266A (open circles) yield of 0.98 mg/L, V43A (filled circles) yield of 0.95 mg/L

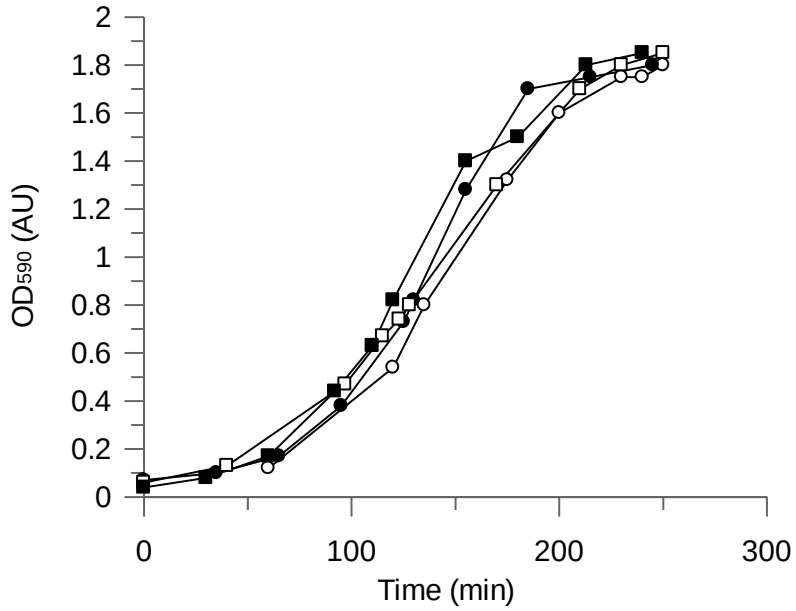


Figure 7 Growth curves for L268A, L44A and L267A and L50A alanine mutants

L268A (open squares) yield of 1.02 mg/L, L44A (filled squares) yield of 1.08 mg/L, L267A (open circles) yield of 0.96 mg/L, L50A (filled circles) yield of 0.97 mg/L

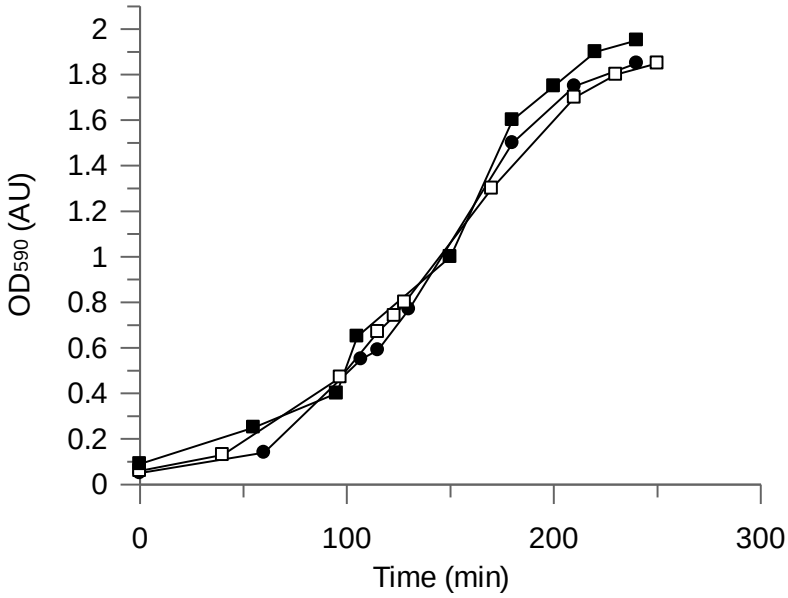


Figure 8 Growth curves for L54A, L273A alanine mutants and cysteine mutant

L54A (open squares) yield of 1.04 mg/L, L273A (filled squares) yield of 0.99 mg/L, (open circles) yield of 0.95 mg/L, Cysteine mutant (filled circles) yield of 0.07 mg/L

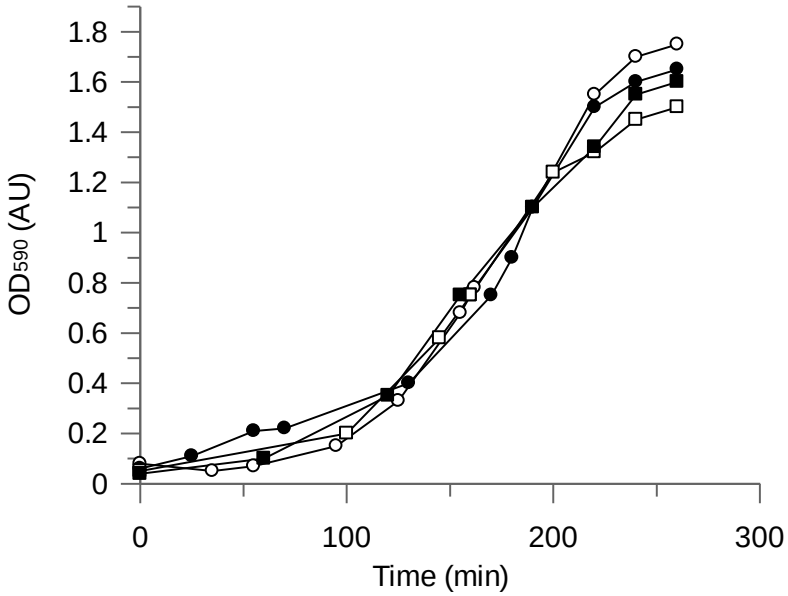


Figure 9 Growth curves for V332, M107W, F148W and Trpless GlpT mutants

V332W (open squares) yield of 0.08 mg/L, M107W (filled squares) yield of 0.08 mg/L, Trpless (open circles) yield of 0.16 mg/L, F148W (filled circles) yield of 0.18 mg/L

Appendix D. PyMOL alanine mutant pairs

The diagrams below illustrate locations of alanine mutant pairs within the GlpT structure. Helix 1 and 7 are displayed in purple and blue, respectively. The mutated residues are labelled in yellow. The positions and possible interactions of alanine mutants are discussed in chapter 4. All structures were rendered in PyMOL software.

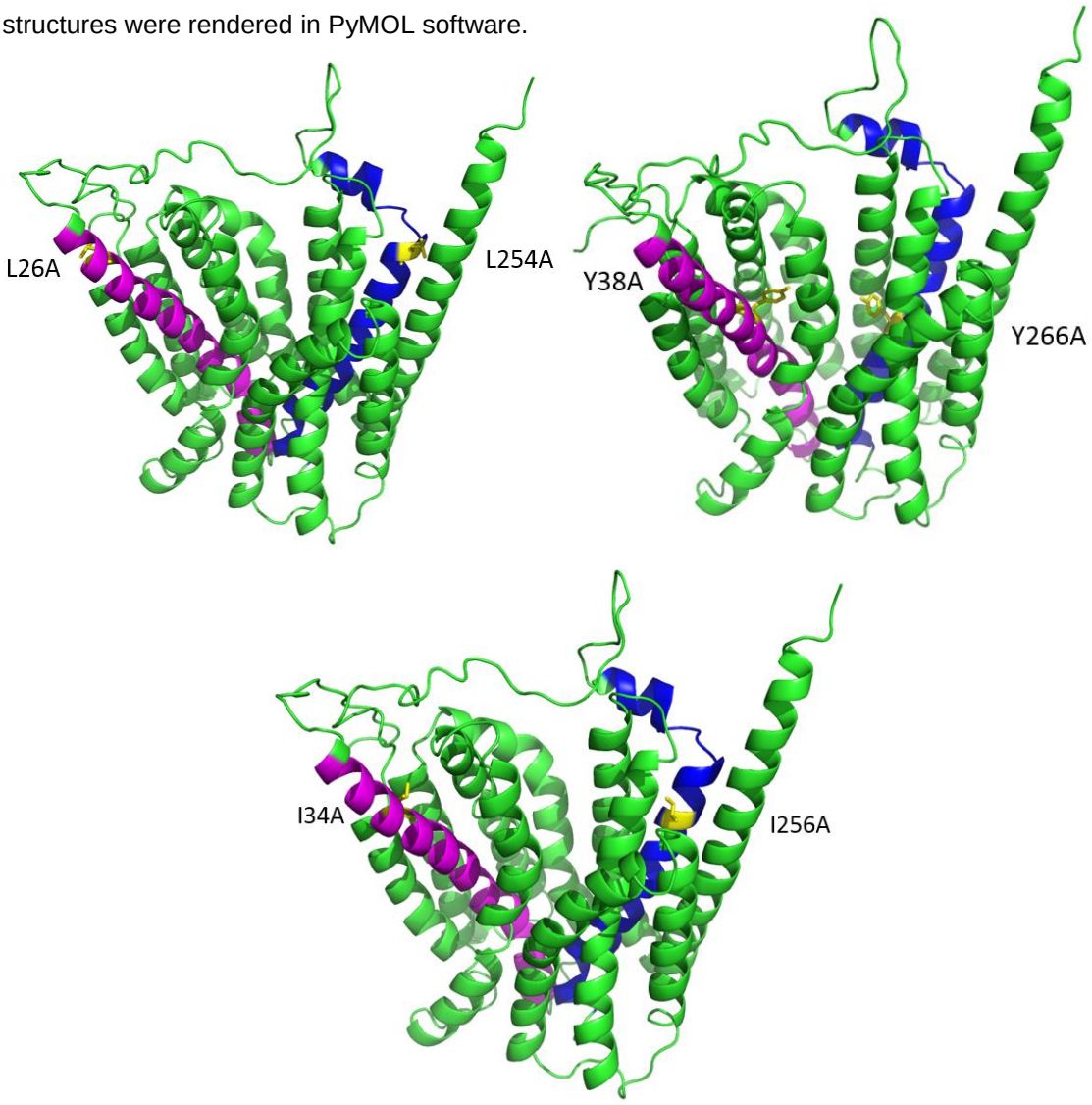


Figure 10 Diagrams L26A-L254A, Y38A-Y266A and I34A-I256A alanine pairs located near cytoplasmic side of the membrane

Helix 1 is represented by purple helix, while helix 7 by blue helix. The mutated residues are indicated in yellow colour.

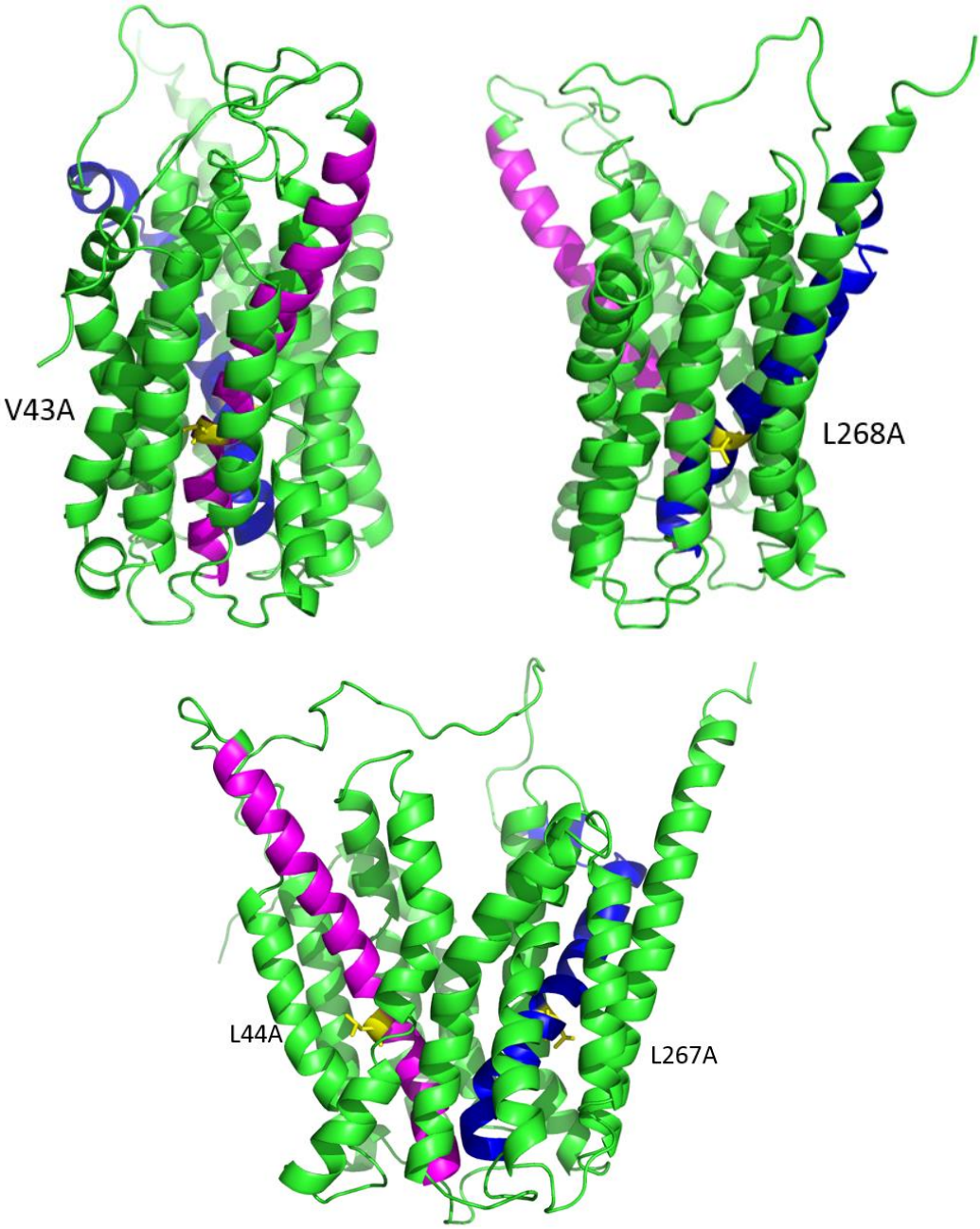


Figure 11 Diagrams for V43A-L268A, L44A-L267A alanine pairs located near protein core

Helix 1 is represented by purple helix, while helix 7 by blue helix. The mutated residues are indicated in yellow colour.

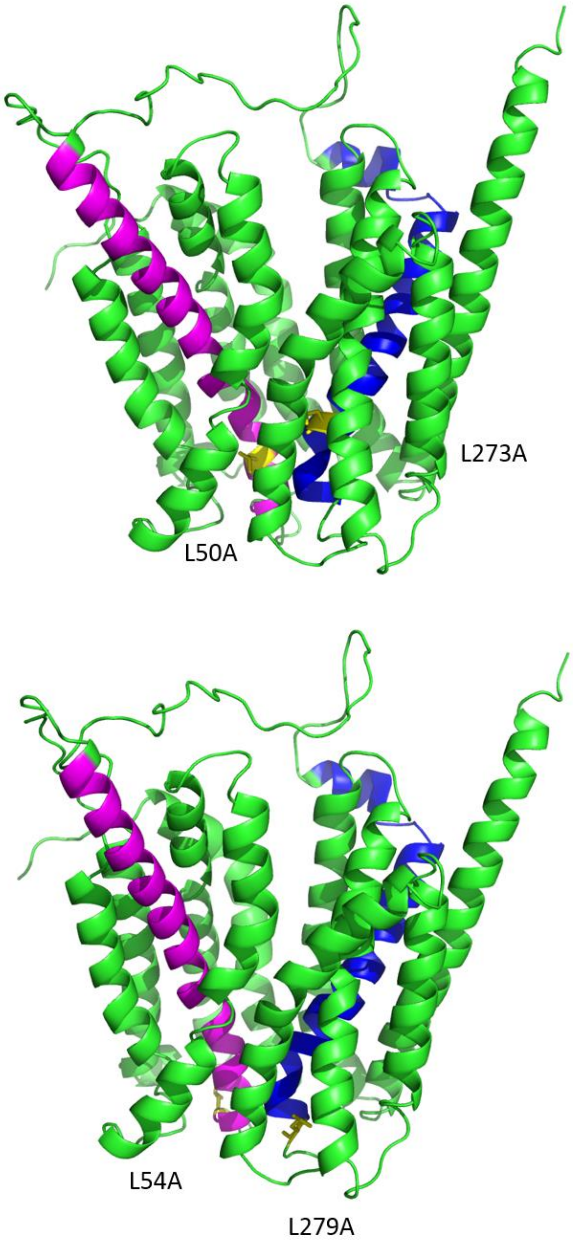


Figure 12 Diagrams for L50A-L273A and L54A-L279A alanine pairs located near periplasmic side of the membrane

Helix 1 is represented by purple helix, while helix 7 by blue helix. The mutated residues are indicated in yellow colour.

Appendix E. References

1 Wallin, E. and Hejine, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial,archaean, and eukaryotic organisms. *Protein Science* **7**, 1029-1039

2 Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**(3): 567-580.

3 Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. and Stewart, P. D. (2014). Membrane protein structure determination - the next generation. *Biochimica et biophysica acta* **1838**(1 Pt A): 78-87.

4 Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**(4096): 223-230.

5 Mackenzie, K. R. (2006). Folding and stability of alpha-helical integral membrane proteins. *Chemical reviews* **106**(5): 1931-1977.

6 Popot, J. L. (2010). Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annual review of biochemistry* **79**: 737-775.

7 Bowie, J. U. (2005). Solving the membrane protein folding problem. *Nature* **438**(7068): 581-589.

8 Lee, S.C., Bennett, B.C., Hong, W.X., Fu. Y., Baker, K.A., Marcoux, J., Robinson, C.V., Ward, A.B., Halpert, J.R., Stevens, R.C., Stout, C.D., Yeager, M.J., Zhang, Q. (2013) Facial amphiphiles for membrane protein studies. *Proceedings of the National Academy of Sciences of the United States of America* **110**1(3):1203-11.

9 Dhakshnamoorthy, B., Raychaudhury, S., Blachowicz, L., Roux, B. (2010) Cation-selective pathway of OmpF porin revealed by anomalous X-ray diffraction. *J.Mol.Biol*. **396**: 293-300.

10 Bowie, J. U. (2011). Membrane protein folding: how important are hydrogen bonds? Current opinion in structural biology **21**(1): 42-49.

11 Killian, J. A. and von Heijne, G. (2000). How proteins adapt to a membrane-water interface. *Trends in biochemical sciences* **25**(9): 429-434.

12 Nick Pace, C., Scholtz, J. M. and Grimsley, G. R. (2014). Forces stabilizing proteins. *FEBS Lett* **588**(14): 2177-2184.

13 Mitchell, D. C. (2012). Progress in understanding the role of lipids in membrane protein folding. *Biochimica et biophysica acta* **1818**(4): 951-956.

14 Booth, P. J. and Curnow, P. (2006). Membrane proteins shape up: understanding in vitro folding. *Current opinion in structural biology* 16**(**4): 480-488.

15 Heijne, G. and Gavel, Y. (1988) Topogenic signals in integral membrane proteins. *European Journal of Biochemistry*, **174,** 671-678.

16 Bogdanov, M., Xie, J. and Dowhan, W. (2009). Lipid-protein interactions drive membrane protein topogenesis in accordance with the positive inside rule. *The Journal of biological chemistry* **284**(15): 9637-9641.

17 Virkki, M. T., Peters, C., Nilsson, D., Sörensen, T., Cristobal, S., Wallner, B., Elofsson., A. (2014) The Positive Inside Rule Is Stronger When Followed by a Transmembrane Helix. *Journal of Molecular Biology*, **426**, 2982-2991

18 Heyden, M., Freites, J. A., Ulmschneider, M. B., White, S. H. and Tobias, D. J. (2012). Assembly and Stability of alpha-Helical Membrane Proteins. *Soft matter* **8**(30): 7742-7752.

19 Karpowich, N. K. and Wang, D. N. (2010). Biophysics: Transporter in the spotlight. *Nature* **465**(7295): 171-172.

20 Huxley, A. F., (2002) Hodgkin and the action potential 1935–1952. *The Journal of Physiology*, **538**, 2.

21 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin,

K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J. and International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

22 Cutting, G. R. (2015). Cystic fibrosis genetics: from molecular understanding to clinical application. *Nature reviews. Genetics* **16**(1): 45-56.

23 Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M., Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences* **25** 331-339.

24 Cheung, M.S., García, A.E., Onuchic, J.N. (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. PNAS, **99**, 685-690.

25 White, S. H. and von Heijne, G. (2005). Transmembrane helices before, during, and after insertion. Current opinion in structural biology **15**(4): 378-386

26 White, S. H. and Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annual review of biophysics and biomolecular structure* **28**: 319-365.

27 Dowhan, W. and Bogdanov, B. (2009) Lipid-Dependent Membrane Protein Topogenesis. *Annual Review of Biochemistry* **78(**1), 515-540.

28 Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature **450**(7172): 1026-1030.

29 Wimley, W. C. and White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature structural biology* **3**(10): 842-848.

30 Wimley, W. C. and White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature structural biology **3**(10): 842-848.

31 Almeida, P. F., Ladokhin, A. S. and White, S. H. (2012). Hydrogen-bond energetics drive helix formation in membrane interfaces. *Biochimica et biophysica acta* **1818**(2): 178-182.

32 Kauko, A., Hedin, L. E., Thebaud, E., Cristobal, S., Elofsson, A. and von Heijne, G. (2010). Repositioning of transmembrane alpha-helices during membrane protein folding. *Journal of molecular biology* **397**(1): 190-201.

33 Almeida, P. F., Ladokhin, A. S. and White, S. H. (2012). Hydrogen-bond energetics drive helix formation in membrane interfaces. *Biochimica et biophysica acta* **1818**(2): 178-182.

34 Kauko, A., Hedin, L. E., Thebaud, E., Cristobal, S., Elofsson, A. and von Heijne, G. (2010). Repositioning of transmembrane alpha-helices during membrane protein folding. Journal of molecular biology **397**(1): 190-201.

35 Skach, W. R. (2009). Cellular mechanisms of membrane protein folding. *Nature structural & molecular biology* **16**(6): 606-612.

36 White, S. H. and von Heijne, G. (2008). How translocons select transmembrane helices. *Annual review of biophysics* **37**: 23-42.

37 Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**(7172): 1026-1030.

38 Li, W., Schulman, S., Boyd, D., Erlandson, K., Beckwith, J. and Rapoport, T. A. (2007). The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Molecular cell* **26**(4): 511-521.

39 Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. and Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**(6969): 36-44.

40 Morgan, D. G., Menetret, J. F., Neuhof, A., Rapoport, T. A. and Akey, C. W. (2002). Structure of the mammalian ribosome-channel complex at 17A resolution. *Journal of molecular biology* **324**(4): 871-886.

41 Cannon, K. S., Or, E., Clemons, W. M., Jr., Shibata, Y. and Rapoport, T. A. (2005). Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *The Journal of cell biology* **169**(2): 219-225.

42 White, S. H. and von Heijne, G. (2005). Transmembrane helices before, during, and after insertion. *Current opinion in structural biology* **15**(4): 378-386.

43 Saraogi, I. and Shan, S. O. (2014). Co-translational protein targeting to the bacterial membrane. *Biochimica et biophysica acta* **1843**(8): 1433-1441.

44 Cross, B. C., Sinning, I., Luirink, J. and High, S. (2009). Delivering proteins for export from the cytosol. *Nature reviews. Molecular cell biology* **10**(4): 255-264.

45 Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**(7024): 377-381.

46 Bogdanov, M., Dowhan, W. and Vitrac, H. (2014). Lipids and topological rules governing membrane protein assembly. *Biochimica et biophysica acta* **1843**(8): 1475-1488.

47 Goder, V., Junne, T. and Spiess, M. (2004). Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Molecular biology of the cell* **15**(3): 1470-1478.

48 Meijberg, W. and Booth, P. J. (2002). The activation energy for insertion of transmembrane alpha-helices is dependent on membrane composition. *Journal of molecular biology* **319**(3): 839-853.

49 Cymer, F., Veerappan, A. and Schneider, D. (2012). Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties. *Biochimica et biophysica acta* **1818**(4): 963-973.

50 Goder, V. and Spiess, M. (2001). Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett* **504**(3): 87-93.

51 Sadlish, H. and Skach, W. R. (2004). Biogenesis of CFTR and other polytopic membrane proteins: new roles for the ribosome-translocon complex. *The Journal of membrane biology* **202**(3): 115-126.

52 Sadlish, H., Pitonzo, D., Johnson, A. E. and Skach, W. R. (2005). Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nature structural & molecular biology* **12**(10): 870-878.

53 Levinthal, C., (1968) Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*. **65**, 44–45.

54 Popot, J. L. and Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**(17): 4031-4037.

55 Kahn, T. W. and Engelman, D. M. (1992). Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry* **31**(26): 6144-6151.

56 Marti, T. (1998). Refolding of bacteriorhodopsin from expressed polypeptide fragments. *The Journal of biological chemistry* **273**(15): 9312-9322.

57 Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., Senes, A. and Popot, J. L. (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett* **555**(1): 122-125.

58 Curnow, P. and Booth, P. J. (2010). The contribution of a covalently bound cofactor to the folding and thermodynamic stability of an integral membrane protein. *Journal of molecular biology* **403**(4): 630-642.

59 Curnow, P. and Booth, P. J. (2009). The transition state for integral membrane protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **106**(3): 773-778.

60 Russ, W. P. and Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *Journal of molecular biology* **296**(3): 911-919.

61 Kleiger, G., Grothe, R., Mallick, P. and Eisenberg, D. (2002). GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* **41**(19): 5990-5997.

62 Takasugi, N., Tomita, T., Hayashi, I., Tsuruoka, M., Niimura, M., Takahashi, Y., Thinakaran, G. and Iwatsubo, T. (2003). The role of presenilin cofactors in the gamma-secretase complex. *Nature* **422**(6930): 438-441.

63 Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**(4860): 1759-1764.

64 Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. and Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nature structural biology* **7**(2): 154-160.

65 Booth, P. J. and Curnow, P. (2009). Folding scene investigation: membrane proteins. *Current opinion in structural biology* **19**(1): 8-13.

66 Zhou, F. X., Merianos, H. J., Brunger, A. T. and Engelman, D. M. (2001). Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America* **98**(5): 2250-2255.

67 Engelman, D. M. (2005). Membranes are more mosaic than fluid. *Nature* **438**(7068): 578-580.

68 Pietzsch, J. (2004) Mind The Membrane. Horizon Symposia

69 Singer, S.J. and Nicolson, G.L., The fluid mosaic model of the structure of cell membranes. *Science*, 1972. **175,** 720-31.

70 Findlay, H. E., Rutherford, N. G., Henderson, P. J. and Booth, P. J. (2010). Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proceedings of the National Academy of Sciences of the United States of America* **107**(43): 18451-18456.

71 van Meer, G., Voelker, D. R. and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. Nature reviews. *Molecular cell biology* **9**(2): 112-124.

72 Wang, X. (2004). Lipid signaling. Current opinion in plant biology **7(**3): 329-336.

73 Booth, P. J. (2000). Unravelling the folding of bacteriorhodopsin. *Biochimica et biophysica acta* **1460**(1): 4-14.

74 Booth, P. J. (2005). Sane in the membrane: designing systems to modulate membrane proteins. *Current opinion in structural biology* **15**(4): 435-440.

75 Booth, P. J., Riley, M. L., Flitsch, S. L., Templer, R. H., Farooq, A., Curran, A. R., Chadborn, N. and Wright, P. (1997). Evidence that bilayer bending rigidity affects membrane protein folding. *Biochemistry* **36**(1): 197-203.

76 Charalambous, K., Miller, D., Curnow, P. and Booth, P. J. (2008). Lipid bilayer composition influences small multidrug transporters. *BMC biochemistry* **9**: 31.

77 Curran, A. R., Templer, R. H. and Booth, P. J. (1999). Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. *Biochemistry* **38**(29): 9328-9336.

78 Bogdanov, M., Sun, J., Kaback, H. R. and Dowhan, W. (1996). A phospholipid acts as a chaperone in assembly of a membrane transport protein. *The Journal of biological chemistry* **271**(20): 11615-11618.

79 Bowie, J. U. (2004). Membrane proteins: a new method enters the fold. *Proceedings of the National Academy of Sciences of the United States of America* **101**(12): 3995-3996.

80 Seddon, A. M., Curnow, P. and Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et biophysica acta* **1666**(1-2): 105-117.

81 Hong, H. and Bowie, J. U. (2011). Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *Journal of the American Chemical Society* **133**(29): 11389-11398.

82 Bowie, J. U. (2001). Stabilizing membrane proteins. *Current opinion in structural biology* **11**(4): 397-402.

83 Bayburt, T. H., Grinkova, Y. V. and Sligar, S. G. (2006). Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs. *Archives of biochemistry and biophysics* **450**(2): 215-222.

84 Dorr, J. M., Scheidelaar, S., Koorengevel, M. C., Dominguez, J. J., Schafer, M., van Walree, C. A. and Killian, J. A. (2016). The styrene-maleic acid copolymer: a versatile tool in membrane research. *European biophysics journal : EBJ* **45**(1): 3-21.

85 Boudker, O. and Verdon, G. (2010). Structural perspectives on secondary active transporters. *Trends in pharmacological sciences* **31**(9): 418-426.

86 Popot, J. L., Berry, E. A., Charvolin, D., Creuzenet, C., Ebel, C., Engelman, D. M., Flotenmeyer, M., Giusti, F., Gohon, Y., Hong, Q., Lakey, J. H., Leonard, K., Shuman, H.

A., Timmins, P., Warschawski, D. E., Zito, F., Zoonens, M., Pucci, B. and Tribet, C. (2003). Amphipols: polymeric surfactants for membrane biology research. *Cellular and molecular life sciences : CMLS* **60**(8): 1559-1574.

87 Sanders, C. R. and Mittendorf, K. F. (2011). Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. *Biochemistry* **50**(37): 7858-7867.

88 Sanders, C. R., Kuhn Hoffmann, A., Gray, D. N., Keyes, M. H. and Ellis, C. D. (2004). French swimwear for membrane proteins. *Chembiochem : a European journal of chemical biology* **5(**4): 423-426.

89 Boudker, O. and Verdon, G. (2010). Structural perspectives on secondary active transporters. *Trends in pharmacological sciences* **31**(9): 418-426.

90 Henderson, P. J. and Baldwin, S. A. (2013). This is about the in and the out. *Nature structural & molecular biology* **20**(6): 654-655.

91 He, M. M. and Kaback, H. R. (1998). In vitro folding of a membrane protein: effect of denaturation and renaturation on substrate binding by the lactose permease of Escherichia coli. *Molecular membrane biology* **15**(1): 15-20.

92 Chou, J. Y. and Mansfield, B. C. (2014). The SLC37 family of sugar-phosphate/phosphate exchangers. Current topics in membranes **73**: 357-382.

93 Lemieux, M. J. (2007). Eukaryotic major facilitator superfamily transporter modeling based on the prokaryotic GlpT crystal structure. *Molecular membrane biology* **24**(5-6): 333-341.

94 Tsigelny, I. F., Greenberg, J., Kouznetsova, V. and Nigam, S. K. (2008). Modeling of glycerol-3-phosphate transporter suggests a potential 'tilt' mechanism involved in its function. *Journal of bioinformatics and computational biology* **6**(5): 885-904.

95 Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. and Saier, M. H., Jr. (2012). The major facilitator superfamily (MFS) revisited. *The FEBS journal* **279**(11): 2022-2035.

96 Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. and Iwata, S. (2003). Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301**(5633): 610-615.

97 Huang, Y., Lemieux, M. J., Song, J., Auer, M. and Wang, D. N. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **301**(5633): 616-620.

98 Yin, Y., He, X., Szewczyk, P., Nguyen, T. and Chang, G. (2006). Structure of the multidrug transporter EmrD from Escherichia coli. *Science* 312(5774): 741-744.

99 Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J. and Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature* **467**(7316): 734-738.

100 Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y. and Yan, N. (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature* **490**(7420): 361-366.

101 Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M. and Yan, N. (2014). Crystal structure of the human glucose transporter GLUT1. *Nature* **510**(7503): 121-125.

102 Kapoor, K., Finer-Moore, J. S., Pedersen, B. P., Caboni, L., Waight, A., Hillig, R. C., Bringmann, P., Heisler, I., Muller, T., Siebeneicher, H. and Stroud, R. M. (2016). Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. *Proceedings of the National Academy of Sciences of the United States of America* **113**(17): 4711-4716.

103 Saier, M. H., Jr., Tran, C. V. and Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research* **34**: 181-186.

104 Mirza, O., Guan, L., Verner, G., Iwata, S., Kaback R, H. (2006) Structural evidence for induced fit and a mechanism for sugar/H+ symport in LacY. *The EMBO Journal,* **25**, 1177-1183.

105 Andreas. P., Spencer, B., Rose, P,W., Bluhm, W.F., Bizon. C., Godzik. A., Bourne, P.E. (2010) Pre-calculated protein structure alignments at the RCSB PDB website Bioinformatics **26,** 2983-2985.

106 Needleman, S.B.. Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. **48**: 443–53.

107 Shindyalov, I.N., Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11**: 739-747.

108 Enomoto, A. and Niwa, T. (2007). Roles of organic anion transporters in the progression of chronic renal failure. *Therapeutic apheresis and dialysis : official peer-reviewed journal of the International Society for Apheresis, the Japanese Society for Apheresis, the Japanese Society for Dialysis Therapy* **11** Suppl 1: S27-31.

109 Smith, D. E., Clémençon, B. & Hediger, M. A. (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol. Aspects Med*. **34**, 323–336.

110 Cura, A. J. & Carruthers, A. Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2**, 863–914 (2012).

111 Almqvist, J., Huang, Y., Hovmoller, S. and Wang, D. N. (2004). Homology modeling of the human microsomal glucose 6-phosphate transporter explains the mutations that cause the glycogen storage disease type Ib. *Biochemistry* **43**(29): 9289-9297.

112 Salas-Burgos, A., Iserovich, P., Zuniga, F., Vera, J. C. and Fischbarg, J. (2004). Predicting the three-dimensional structure of the human facilitative glucose transporter glut1 by a novel evolutionary homology strategy: insights on the molecular mechanism of substrate migration, and binding sites for glucose and inhibitory molecules. *Biophysical journal* **87**(5): 2990-2999.

113 Matthews, C. R. (1993). Pathways of protein folding. *Annual review of biochemistry* **62**: 653-683.

114 Radestock, S. and Forrest, L. R. (2011). The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of molecular biology* **407**(5): 698-715.

115 Quistgaard, E. M., Low, C., Moberg, P., Tresaugues, L. and Nordlund, P. (2013). Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature structural & molecular biology* **20**(6): 766-768.

116 Smirnova, I., Kasho, V. and Kaback, H. R. (2011). Lactose permease and the alternating access mechanism. *Biochemistry* **50**(45): 9684-9693.

117 Law, C. J., Almqvist, J., Bernstein, A., Goetz, R. M., Huang, Y., Soudant, C., Laaksonen, A., Hovmoller, S. and Wang, D. N. (2008). Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *Journal of molecular biology* **378**(4): 828-839.

118 Newstead, S., Drew, D., Cameron, A. D., Postis, V. L., Xia, X., Fowler, P. W., Ingram, J. C., Carpenter, E. P., Sansom, M. S., McPherson, M. J., Baldwin, S. A. and Iwata, S. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO journal* **30**(2): 417-426.

119 Law, C. J., Enkavi, G., Wang, D. N. and Tajkhorshid, E. (2009). Structural basis of substrate selectivity in the glycerol-3-phosphate: phosphate antiporter GlpT. *Biophysical journal* **97**(5): 1346-1353.

120 Fann, M. C., Busch, A. and Maloney, P. C. (2003). Functional characterization of cysteine residues in GlpT, the glycerol 3-phosphate transporter of Escherichia coli. *J Bacteriol* **185**(13): 3863-3870.

121 Zhou, Y., Guan, L., Freites, J. A. and Kaback, H. R. (2008). Opening and closing of the periplasmic gate in lactose permease. Proceedings of the National *Academy of Sciences of the United States of America* **105**(10): 3774-3778.

122 Enkavi, G. and Tajkhorshid, E. (2010). Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT. *Biochemistry* **49**(6): 1105-1114.

123 Smirnova, I., Kasho, V., Choe, J. Y., Altenbach, C., Hubbell, W. L. and Kaback, H. R. (2007). Sugar binding induces an outward facing conformation of LacY. *Proceedings of the National Academy of Sciences of the United States of America* **104**(42): 16504-16509.

124 Findlay, H. E. and Booth, P. J. (2006). The biological significance of lipid-protein interactions. *Journal of physics. Condensed matter : an Institute of Physics journal* **18**(28): S1281-1291.

125 Harris, N. J., Findlay, H. E., Simms, J., Liu, X. and Booth, P. J. (2014). Relative domain folding and stability of a membrane transport protein. *Journal of molecular biology* **426**(8): 1812-1825.

126 Curnow, P., Di Bartolo, N. D., Moreton, K. M., Ajoje, O. O., Saggese, N. P. and Booth, P. J. (2011). Stable folding core in the folding 7-helical integral membrane protein. *Proceedings of the National Academy of Sciences of the United States of America* **108**(34): 14133-14138.

127 Cymer, F. and von Heijne, G. (2013). Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proceedings of the National Academy of Sciences of the United States of America* **110**(36): 14640-14645.

128 D'Rozario, R. S. and Sansom, M. S. (2008). Helix dynamics in a membrane transport protein: comparative simulations of the glycerol-3-phosphate transporter and its constituent helices. *Molecular membrane biology* **25**(6-7): 571-583.

129 Auer, M., Kim, M. J., Lemieux, M. J., Villa, A., Song, J., Li, X. D. and Wang, D. N. (2001). High-yield expression and functional analysis of Escherichia coli glycerol-3-phosphate transporter. *Biochemistry* **40**(22): 6628-6635.

130 Madej, M. G., Dang, S., Yan, N. and Kaback, H. R. (2013). Evolutionary mix-and-match with MFS transporters. *Proceedings of the National Academy of Sciences of the United States of America* **110**(15): 5870-5874.

131 Moradi, M., Enkavi, G. and Tajkhorshid, E. (2015). Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nature communications* **6**: 8393.

132 Shen, W., Wei, Y., Dauk, M., Tan, Y., Taylor, D. C., Selvaraj, G. and Zou, J. (2006). Involvement of a glycerol-3-phosphate dehydrogenase in modulating the NADH/NAD+ ratio provides evidence of a mitochondrial glycerol-3-phosphate shuttle in Arabidopsis. *The Plant cell* **18**(2): 422-441.

133 Bartoloni, L., Wattenhofer, M., Kudoh, J., Berry, A., Shibuya, K., Kawasaki, K., Wang, J., Asakawa, S., Talior, I., Bonne-Tamir, B., Rossier, C., Michaud, J., McCabe, E. R., Minoshima, S., Shimizu, N., Scott, H. S. and Antonarakis, S. E. (2000). Cloning and characterization of a putative human glycerol 3-phosphate permease gene (SLC37A1 or G3PP) on 21q22.3: mutation analysis in two candidate phenotypes, DFNB10 and a glycerol kinase deficiency. *Genomics* **70**(2): 190-200.

134 Castaneda-Garcia, A., Rodriguez-Rojas, A., Guelfo, J. R. and Blazquez, J. (2009). The glycerol-3-phosphate permease GlpT is the only fosfomycin transporter in Pseudomonas aeruginosa. *J Bacteriol* **191**(22): 6968-6974.

135 Santoro, A., Cappello, A. R., Madeo, M., Martello, E., Iacopetta, D. and Dolce, V. (2011). Interaction of fosfomycin with the glycerol 3-phosphate transporter of Escherichia coli. *Biochimica et biophysica acta* **1810**(12): 1323-1329.

136 Compton, E. L., Farmer, N. A., Lorch, M., Mason, J. M., Moreton, K. M. and Booth, P. J. (2006). Kinetics of an individual transmembrane helix during bacteriorhodopsin folding. *Journal of molecular biology* **357**(1): 325-338.

137 Booth, P. J. and Clarke, J. (2010). Membrane protein folding makes the transition. *Proceedings of the National Academy of Sciences of the United States of America* **107**(9): 3947-3948.

138 Fersht, A. R. and Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states*. Proceedings of the National Academy of Sciences of the United States of America* **101**(21): 7976-7981.

139 Booth, P. J. (2012). A successful change of circumstance: a transition state for membrane protein folding. *Current opinion in structural biology* **22**(4): 469-475.

140 Morrison, K. L. and Weiss, G. A. (2001). Combinatorial alanine-scanning. *Current opinion in chemical biology* **5**(3): 302-307.

141 Edelheit, O., Hanukoglu, I., Dascal, N. and Hanukoglu, A. (2011). Identification of the roles of conserved charged residues in the extracellular domain of an epithelial sodium channel (ENaC) subunit by alanine mutagenesis. *American journal of physiology. Renal physiology* **300**(4): 887-897.

142 Lees, J. G., Smith, B. R., Wien, F., Miles, A. J. and Wallace, B. A. (2004). CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Analytical biochemistry* **332**(2): 285-289.

143 Kompella, S. N., Hung, A., Clark, R. J., Mari, F. and Adams, D. J. (2015). Alanine scan of alpha-conotoxin RegIIA reveals a selective alpha3beta4 nicotinic acetylcholine receptor antagonist. *The Journal of biological chemistry* **290**(2): 1039-1048.

144 Baker, S. H., Frederick, D. L., Bloecher, A. and Tatchell, K. (1997). Alanine-scanning mutagenesis of protein phosphatase type 1 in the yeast Saccharomyces cerevisiae. *Genetics* **145**(3): 615-626.

145 Bian, F., Yue, S., Peng, Z., Zhang, X., Chen, G., Yu, J., Xuan, N. and Bi, Y. (2015). A comprehensive alanine-scanning mutagenesis study reveals roles for salt bridges in the structure and activity of Pseudomonas aeruginosa elastase. *PloS one* **10**(3): 12-18.

146 Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook:* 571-607.

147 Abdul-Gader, A., Miles, A. J. and Wallace, B. A. (2011). A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* **27**(12): 1630-1636.

148 Compton, L.A. and Johnson, W.C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **155**: 155-167.

149 Abdul-Gader, A., Miles, A.J., Wallace, B.A. (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics,* **27**: 1630-1636.

150 Lees, J.G., Miles, A.J., Wien, F., and Wallace, B.A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics,* **22:** 1955-1962.

151 Markwell, M.A.K, Haas, S.M, L.L. Bieber, N.E. (1978) A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples, *Anal. Biochem*, **87**, 206-210.

152 Harris, N. J. and Booth, P. J. (2012). Folding and stability of membrane transport proteins in vitro. *Biochimica et biophysica acta* **1818**(4): 1055-1066.

153 Lau, F. W. and Bowie, J. U. (1997). A method for assessing the stability of a membrane protein. *Biochemistry* **36**(19): 5884-5892.

154 Tanford, C. (1970) Protein denaturation. Theoretical models for the mechanism of denaturation. *Adv Protein Chem*, **24:** 1-95.

155 Myers, J. K., Pace, C. N. and Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein science : a publication of the Protein Society* **4**(10): 2138-2148.

156 O'Brien, E. P., Dima, R. I., Brooks, B. and Thirumalai, D. (2007). Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *Journal of the American Chemical Society* **129**(23): 7346-7353.

157 Pace, C.N., (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curve. *Methods Enzymol*. **131:** 266-280.

158 Kelly, S. M. and Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Current protein & peptide science* **1**(4): 349-384.

159 Bretscher, M. S. and Munro, S. (1993). Cholesterol and the Golgi apparatus. *Science* **261**(5126): 1280-1281.

160 Faham S (2004). Side-chain contributions to membrane protein stability and structure. *J Mol Biol*. **335,** 297–305.

161 Eftink, M. R. (1994). The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophysical journal* **66**(2 Pt 1): 482-501.

162 Yin, J. and Jing, G. (2000). Tryptophan 140 is important, but serine 141 Is essential for the formation of the integrated conformation of staphylococcal nuclease. *Journal of biochemistry* **128**(1): 113-119.

163 Daggett, V. and Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends in biochemical sciences* **28**(1): 18-25.

164 Royer, C. A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chemical reviews* **106**(5): 1769-1784.

165 Ladokhin, A. S. (2006). Fluorescence Spectroscopy in Peptide and Protein Analysis. In Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd.

166 Scholes, G. D. and Ghiggino, K. P. (1994). Electronic Interactions and Interchromophore Excitation Transfer. *The Journal of Physical Chemistry* **98**(17): 4580-4590.

167 Kunzelmann, S. and Webb, M. R. (2010). A fluorescent, reagentless biosensor for ADP based on tetramethylrhodamine-labeled ParM. *ACS chemical biology* **5**(4): 415-425.

168 Perieteanu, A. A., Sweeting, B. and Dawson, J. F. (2008). The real-time monitoring of the thermal unfolding of tetramethylrhodamine-labeled actin. *Biochemistry* **47**(36): 9688-9696.

169 Okoh, M. P., Hunter, J. L., Corrie, J. E. T. and Webb, M. R. (2006). A Biosensor for Inorganic Phosphate Using a Rhodamine-Labeled Phosphate Binding Protein. *Biochemistry* **45**(49): 14764-14771.

170 Hamman, B. D., Oleinikov, A. V., Jokhadze, G. G., Bochkariov, D. E., Traut, R. R. and Jameson, D. M. (1996). Tetramethylrhodamine dimer formation as a spectroscopic probe of the conformation of Escherichia coli ribosomal protein L7/L12 dimers. *The Journal of biological chemistry* **271**(13): 7568-7573.

171 Eftink, M. R. (2000). Use of fluorescence spectroscopy as thermodynamics tool. *Methods in enzymology* **323**: 459-473.

172 van de Weert, M. (2010). Fluorescence quenching to study protein-ligand binding: common errors. *Journal of fluorescence* **20**(2): 625-629.

173 van de Weert, M. and Stella, L. (2011). Fluorescence quenching and ligand binding: A critical discussion of a popular methodology. *Journal of Molecular Structure* **998**(1–3): 144-150.

174 Marme, N., Knemeyer, J. P., Sauer, M. and Wolfrum, J. (2003). Inter- and intramolecular fluorescence quenching of organic dyes by tryptophan. *Bioconjugate chemistry* **14**(6): 1133-1139.

175 Grama, L., Somogyi, B. and Kellermayer, M. S. (2001). Global configuration of single titin molecules observed through chain-associated rhodamine dimers. *Proceedings of the National Academy of Sciences of the United States of America* **98**(25): 14362-14367.

176 Mackenzie, K. R. (2006). Folding and stability of alpha-helical integral membrane proteins. *Chemical reviews* **106**(5): 1931-1977.

177 Otzen, D.E. (2011). Mapping the folding pathway of the transmembrane protein DsbB by protein engineering. *Protein Engineering Design & Selection* **24**: 139-149.

178 Curnow, P. and Booth, P. J. (2007). Combined kinetic and thermodynamic analysis of alpha-helical membrane protein unfolding. *Proceedings of the National Academy of Sciences of the United States of America* **104**(48): 18970-18975.

179 Van Petegem, F., Duderstadt, K. E., Clark, K. A., Wang, M. and Minor, D. L., Jr. (2008). Alanine-scanning mutagenesis defines a conserved energetic hotspot in the CaValpha1 AID-CaVbeta interaction site that is critical for channel modulation. *Structure* **16**(2): 280-294.

180 Matouschek, A. and Fersht, A. R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **90**(16): 7814-7818.

181 Sanchez, K. M., Kang, G., Wu, B. and Kim, J. E. (2011). Tryptophan-lipid interactions in membrane protein folding probed by ultraviolet resonance Raman and fluorescence spectroscopy. *Biophysical journal* **100**(9): 2121-2130.